LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN
numero 65
gennaio 1974
anno VII
volume XII

Le collisioni elettrone-positone

Quando un elettrone e la sua antiparticella entrano in collisione, essi possono annichilarsi e dare origine a radiazioni o ad altre particelle. Tali processi vengono comunemente studiati con fasci di elettroni e positoni

di Alan M. Litke e Richard Wilson

Cosa accade quando una veloce particella di materia ordinaria urta frontalmente una veloce particella di antimateria? La questione è comunemente esaminata in parecchi laboratori di fisica delle alte energie sparsi nel mondo, dove i ricercatori sono impegnati ad analizzare i vari tipi di interazioni che possono aver luogo tra due fasci di particelle in direzione opposta, uno costituito da elettroni accelerati e l'altro da positoni, o antielettroni, accelerati.

La risposta, basata sugli esperimenti già portati a termine, è uno sconcertante miscuglio di situazioni previste e di imprevisti. Finora gli esperimenti hanno confermato la validità alle alte energie della teoria fondamentale dei fenomeni elettromagnetici che va sotto il nome di elettrodinamica quantistica; essi hanno inoltre rivelato che le collisioni nelle quali elettroni e positoni si annichilano per produrre particelle nucleari a interazione forte si verificano piú frequentemente di quanto previsto. Quest'ultima scoperta ha importanti conseguenze sulla nostra comprensione delle particelle elementari.

L'elettrone è il ben noto costituente degli atomi che ruota attorno ai molto piú pesanti nuclei atomici; esso ha una carica elettrica negativa, una massa di $9,1 \times 10^{-28}$ grammi e uno spin, o momento angolare, di 1/2 (in unità $h/2\pi$, dove h, la grandezza fisica fondamentale nota come costante di Planck, è uguale a $6,6 \times 10^{-27}$ erg secondo). L'elettrone e il positone sono uno l'antiparticella dell'altro; hanno la stessa massa e lo stesso spin a carica elettrica uguale, ma di segno opposto. Quindi, quando un positone viene immerso in un campo elettrico o si muove in un campo magnetico, esso « sente » una forza che è esattamente opposta a quella sentita dall'elettrone.

L'esistenza di un equivalente antimateriale dell'elettrone fu postulata per la prima volta 45 anni fa da P.A.M. Dirac. In quel tempo Dirac stava cercando di impostare una teoria quantomeccanica dell'elettrone che potesse soddisfare la teoria della relatività speciale di Einstein. La migliore formulazione precedente era inadeguata poiché non soddisfaceva i requisiti di simmetria dello spazio e del tempo della teoria speciale. Dirac scoprí che la sua soluzione non solo era in grado di descrivere con successo molti fenomeni noti, ma prevedeva per l'elettrone la esistenza di stati a energia negativa. All'inizio questi ipotetici stati energetici furono ignorati, ma in seguito Dirac si accorse che essi potevano essere spiegati supponendo che in condizioni normali tali stati siano occupati da elettroni.

Se uno di tali stati a energia negativa non fosse occupato, allora, secondo le equazioni di Dirac dovrebbe crearsi una « buca »; tale buca dovrebbe comportarsi rispetto all'elettrone come una identica particella con la sola differenza che dovrebbe avere carica positiva. Dirac chiamò queste ipotetiche buche, simili a particelle, col nome di positoni. Se si deve produrre un positone, affermò, si deve spostare un elettrone dal « mare » di energia negativa, lasciando al suo posto una buca. Ne segue che un positone deve essere prodotto sempre in coppia con un elettrone.

Subito dopo Carl D. Anderson riuscí a produrre coppie elettrone-positone. Quando egli bombardò un setto di piombo di una camera a nebbia con raggi gamma di una sorgente radioattiva, si poterono osservare coppie di tracce che nascevano dallo stesso punto nel piombo. In presenza di un campo magnetico le tracce curvavano in versi opposti. Il processo rappresenta chiaramente una trasformazione di energia radiante in materia. L'energia del campo elettromagnetico del raggio gamma era stata trasformata in energia di massa e in energia cinetica di una coppia elettrone-positone.

Oggi si dà per scontato che ogni particella abbia una corrispondente antiparticella (anche se alcune particelle elettricamente neutre come il fotone e il pione neutro sono le antiparticelle di se stesse) e che con un raggio gamma di energia sufficiente si possa produrre qualsiasi coppia particella-antiparticella. Viceversa, qualsiasi particella, incontrando la propria antiparticella, può subire una transizione verso stati a energia negativa, con l'annichi-

lazione di entrambe le particelle e la produzione di energia radiante.

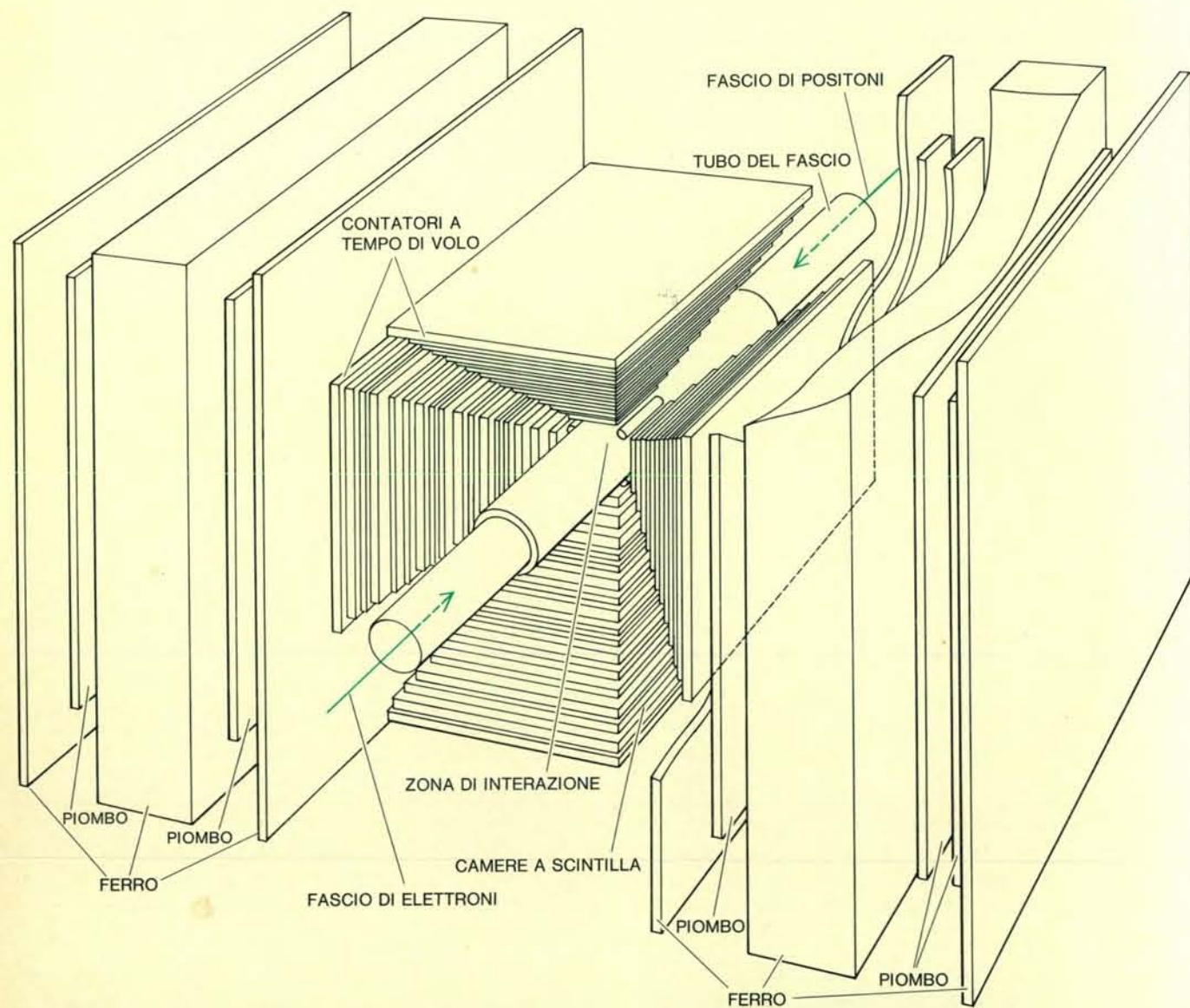
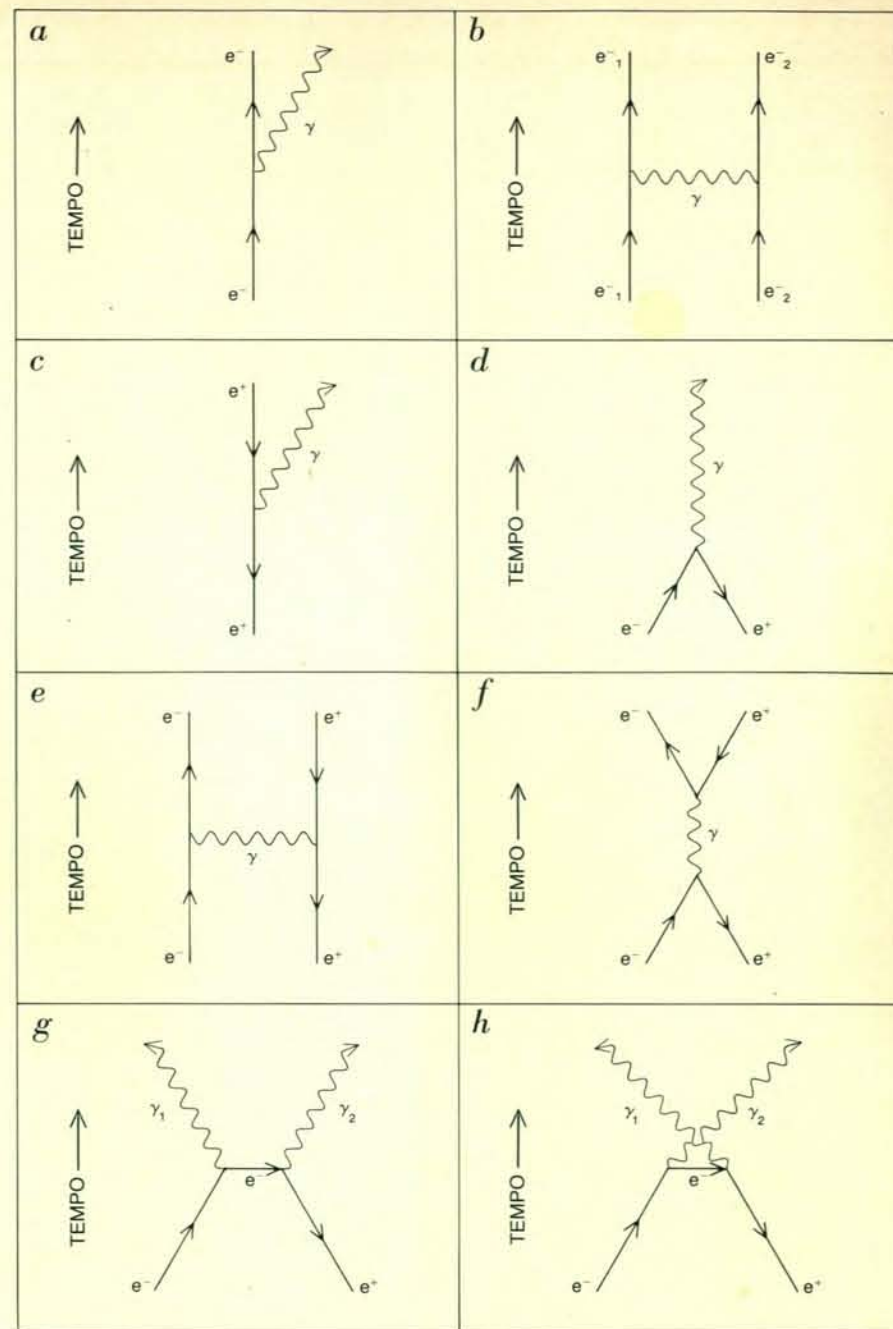
Negli esperimenti che descriveremo i fisici hanno studiato la mutua annichilazione dell'elettrone e del positone. Se le particelle iniziali hanno energia sufficiente, l'annichilazione può dare origine sia a coppie particella-antiparticella sia a energia radiante. Questi esperimenti costituiscono una sorta di moderna alchimia in cui una coppia elettrone-positone può essere convertita in un altro tipo di coppia particella-antiparticella, quale un protone e un antiprotone. Questo tipo di reazione non è una sorpresa per il fisico: è solo una manifestazione dell'intercambiabilità della materia con l'energia, idea che fu avanzata per la prima volta da Albert Einstein nel 1905.

Tali esperimenti richiedono che un fascio di elettroni sia indirizzato contro un fascio di positoni in modo che le particelle possano urtarsi ad alta energia; essi non poterono essere eseguiti fino a quando non vennero risolti certi problemi di carattere tecnico. Uno dei primi problemi incontrati fu quello di come procurarsi un numero sufficiente di positoni. È abbastanza facile produrre fasci di elettroni (lo conferma il grande numero di impianti televisivi esistenti al mondo), poiché evaporano istantaneamente da un filamento di tungsteno riscaldato, ma i positoni, almeno sul nostro pianeta, sono pochi e molto distanti tra loro. Alcuni positoni vengono prodotti nelle collisioni dei raggi cosmici, mentre altri sono emessi da certi nuclei radioattivi, ma tali sorgenti non sono sufficientemente intense da generare un fascio di positoni.

Il problema della produzione dei positoni può essere risolto con il seguente stratagemma. Se si lascia cadere un fascio di elettroni di alta energia su un bersaglio di tungsteno, all'interno del bersaglio gli elettroni, accelerati dal campo elettrico dei nuclei bersaglio irradiano energia elettromagnetica, costituita in parte da raggi gamma. I raggi gamma, mentre attraversano il bersaglio, possono dare origine a coppie elettrone-positone, le quali a loro volta possono produrre altri raggi gamma. Il risultato è uno sciame elettromagnetico, un fenomeno in cascata nel quale vengono prodotti molti elettroni, positoni e raggi gamma. L'ultimo passo è l'eliminazione degli elettroni e la cattura dei positoni uscenti dal bersaglio. Poiché le particelle hanno carica opposta, possono essere separate con un sistema di lenti magnetiche.

In linea di principio il fascio di positoni risultante potrebbe ora essere lasciato cadere su qualsiasi bersaglio contenente elettroni; dopo tutto vi sono circa 10^{24} elettroni in un centimetro cubo di materia solida. In questo caso, però, non solo si avrebbe un effetto di fondo dovuto alle collisioni dei positoni con i nuclei atomici del bersaglio, ma anche l'energia efficace delle collisioni positone-elettrone risulterebbe troppo bassa.

Fino al 1965 tutti gli esperimenti di fisica delle alte energie erano eseguiti bombardando un bersaglio fisso con un fascio di particelle energetiche. Questa tecnica, però, provoca uno spreco di energia come si può dimostrare facilmente con la seguente analogia. Lo scontro tra due automobili che viaggiano nello stesso senso non è particolarmente energetico e ne può risultare un danno modesto, mentre lo scontro tra un'automobile in moto e una ferma è piú energetico e può avere piú serie conseguenze; uno scontro frontale tra due automobili in moto è altamente energetico ed è solitamente catastrofico. L'energia « utile » ricavata da ciascuno dei tre scontri (cioè, la quantità di energia di massa e di energia cinetica dei veicoli urtantisi che può essere trasformata in altre fonti di energia, quali il calore) si può valutare calcolando l'energia totale iniziale misurata da un osservatore che vede le due automobili andare una incontro all'altra con quantità di moto uguali e opposte. Un siffatto osservatore si dice che è solidale col sistema di riferimento del centro di massa.

Nel caso di una collisione tra una particella ferma e una di alta energia l'energia utile può essere solo una modesta frazione dell'energia iniziale (una frazione molto piú piccola di quanto suggerisce la precedente analogia). Ciò deriva dalle leggi relativistiche di conservazione dell'energia e della quantità di moto. Se due particelle con quantità di moto uguali e opposte si urtano frontalmente l'energia utile è invece uguale all'energia iniziale. Per esempio, un positone da 2500 milioni di elettronvolt (o 2,5 GeV) che urta un elettrone stazionario produce un'energia utile nel centro di massa pari al

Il rivelatore a sovrappasso in linea (BOLD) del Cambridge Electron Accelerator, fu usato dagli autori di questo articolo e dai loro colleghi per studiare ciò che accade nella collisione di particelle di materia e di antimateria. Un fascio di elettroni a un'energia di 2000 milioni di elettronvolt (MeV) entrava nel tubo del fascio da sinistra, mentre un fascio di positoni della stessa energia entrava da destra. La zona di interazione era circondata da una apparecchiatura di rivelazione costituita da camere a scintilla, da contatori a scintillazione e da lastre metalliche disposte in modo da formare quattro quadranti identici.

I diagrammi di Feynman, ideati da Richard P. Feynman dell'Istituto di tecnologia della California, mostrano le varie interazioni di particelle elementari, in questo caso elettroni, positoni e fotoni (quanti di radiazione). L'emissione di luce da parte di un atomo o la produzione di radioonde da parte degli elettroni di un'antenna sono basati su una singola reazione (*a*) nella quale un elettrone (e^-) emette un fotone (γ). Questa reazione non potrà mai essere osservata, poiché non si conservano contemporaneamente l'energia e la quantità di moto. Essa è chiamata perciò reazione virtuale. È però possibile provocare una successione di reazioni virtuali per ottenere processi osservabili nei quali si conservano sia l'energia che la quantità di moto. Per esempio, il fotone emesso da un elettrone (e_1) può essere assorbito da un secondo elettrone (e_2). Il processo complessivo (*b*) rappresentato da questa reazione è la diffusione di un elettrone da parte di un altro mediante forze elettromagnetiche; la reazione complessiva avviene con conservazione sia dell'energia che della quantità di moto. Il positone (e^+) può prender parte a interazioni di questo tipo e può anch'esso emettere un fotone virtuale (*c*). In tutti i diagrammi il tempo cresce verso l'alto; il positone differisce dall'elettrone solo per il fatto che la sua freccia direzionale è rivolta all'indietro. Un elettrone e un positone possono annichilarsi per dare origine a un fotone (*d*). Questi processi virtuali possono essere combinati per ricavare una rappresentazione della diffusione elastica tra un positone e un elettrone (*e* ed *f*). In un caso (*e*) l'elettrone emette un fotone virtuale che viene assorbito dal positone. Nell'altro (*f*) l'elettrone e il positone si annichilano per dare origine a un fotone virtuale che viene poi trasformato in una coppia elettrone-positone. Un'interazione finale è la conversione di una coppia positone-elettrone in due fotoni (*g* e *h*); in tal caso viene scambiato un elettrone virtuale. Il motivo per cui vi sono due diagrammi di Feynman per questo processo è che è impossibile stabilire quale raggio gamma è stato emesso.

solo 2 % dell'energia iniziale, lasciando a disposizione solo 50 MeV per la produzione di nuove particelle. D'altra parte, quando positoni da 2500 MeV urtano frontalmente elettroni in un fascio da 2500 MeV, per la produzione di nuove particelle è disponibile l'energia totale delle particelle dei due fasci: 5000 MeV.

Il sistema piú efficace per ottenere alte energie nel sistema di riferimento del centro di massa è dunque quello di provocare collisioni frontali tra i due fasci; ciò però conduce a un'altra maggiore difficoltà tecnica: come si può fare per ottenere collisioni tra le particelle dei fasci a un ritmo sufficientemente alto da permetterci di studiare le rarissime reazioni di annichilazione che ci interessano?

Qual è la frequenza di questi eventi e quale frequenza di collisione è necessaria per questo tipo di analisi? La probabilità di reazione tra due particelle qualsiasi si può esprimere in funzione della sezione d'urto della reazione, la quale può a sua volta essere interpretata come quell'area in prossimità di una delle particelle (considerata come bersaglio) all'interno della quale le forze responsabili della reazione vengono sentite dall'altra particella (considerata come proiettile). Le sezioni d'urto delle reazioni che vogliamo analizzare sono davvero piccole: esse sono comprese tra 10^{-30} e 10^{-34} centimetri quadrati.

Un processo prodotto da collisione tra fasci ha una frequenza uguale alla sezione d'urto di reazione moltiplicata per un fattore chiamato luminosità. La luminosità è una caratteristica della macchina nella quale ha luogo la collisione tra particelle ed è una misura della frequenza con cui una particella di un fascio passa vicina a una particella dell'altro.

Per poter disporre di una luminosità abbastanza alta da poter studiare i rari eventi di annichilazione conviene dare alle particelle di un fascio la possibilità di entrare in collisione ripetutamente con quelle dell'altro. Ciò si può ottenere immagazzinando e accumulando le particelle in un dispositivo circolare chiamato anello di accumulazione. Nell'anello vengono iniettati successivi impulsi di particelle i quali vengono fatti circolare molte volte in un'orbita chiusa. Due fasci possono circolare in anelli separati che si intersecano in uno o piú punti; nel caso di particelle con cariche elettriche opposte, i fasci possono circolare in senso opposto nello stesso anello. Le collisioni tra le particelle di due fasci opposti hanno luogo in punti chiamati zone di interazione, che devono trovarsi nei punti dove le orbite si intersecano. Anelli di accumulazione per elettroni-positoni sono in funzione o in costruzione a Novosibirsk, in URSS, ad Amburgo, in Germania, a Frascati, in Italia, a Orsay, in Francia e presso l'Università di Stanford negli USA. Il primo anello di accumulazione per elettroni-positoni degli USA, il Cambridge Electron Accelerator di Cambridge, è stato chiuso nel giugno 1973.

I principali limiti alla luminosità ottenibile in un anello di accumulazione sono costituiti dalle instabilità che originano dall'interazione dei due fasci; in un anello di accumulazione solo alcune orbite sono stabili: quando i due fasci opposti si incrociano in una zona di interazione, ognuno di essi esercita sull'altro potenti forze elettriche e magnetiche che ne sospingono l'orbita verso una regione di instabilità.

Un metodo per migliorare la luminosità di un anello di accumulazione è stato proposto nel 1965 da K.W. Robinson e G.A. Voss del gruppo di Cambridge. Essi proposero l'impiego di un sistema di lenti magnetiche a focalizzazione eccezionalmente forte per concentrare i fasci in una piccola area trasversale nelle zone di interazione. Quando i fasci sono confinati a una piccola area trasversale, ogni particella di un fascio si trova necessariamente vicina alle particelle dell'altro e ciò aumenta la probabilità di collisione e quindi la luminosità.

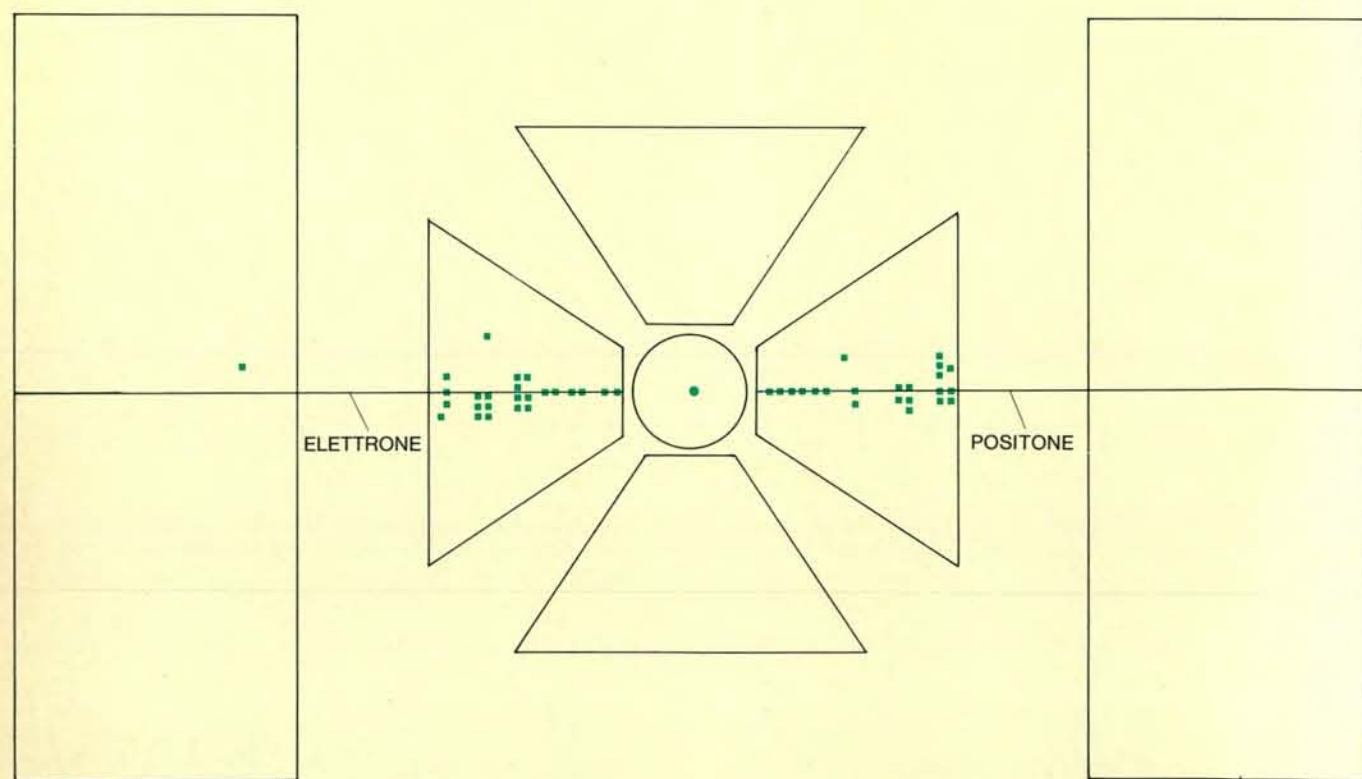
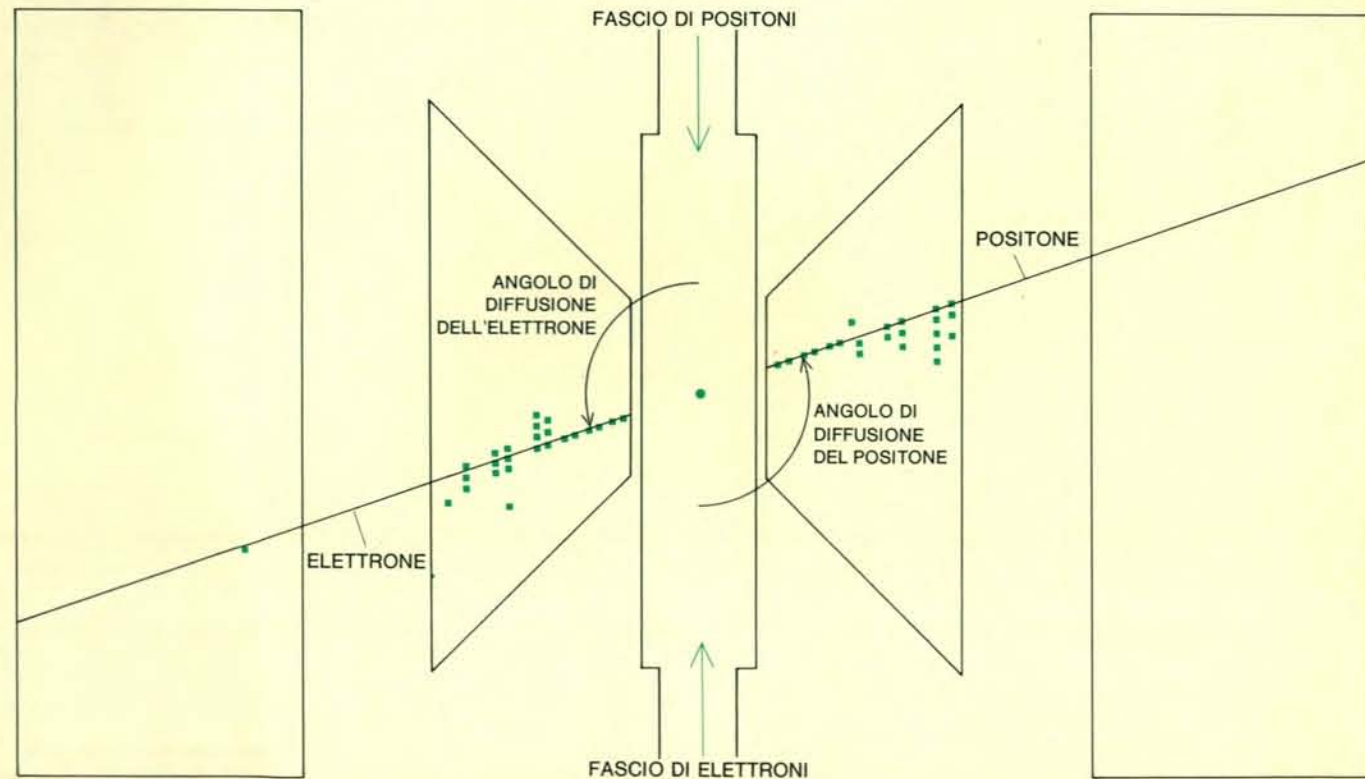
Sfortunatamente gli effetti defocalizzanti di un fascio sull'altro aumentano al diminuire delle dimensioni del fascio. Le forze defocalizzanti sono di natura elettromagnetica e seguono quindi la legge di Coulomb, la quale afferma che la forza è inversamente proporzionale al quadrato della distanza tra le cariche. Se la diminuzione delle dimensioni del fascio è seguita, o causata, da un adeguato aumento di focalizzazione si può riuscire a compensare tale effetto defocalizzante. La tecnica della focalizzazione forte ci permette di ottenere perciò un minore « inviluppo del fascio ». Questa tecnica è stata impiegata per la prima volta nel 1970 al Cambridge Electron Accelerator ed è stata usata recentemente allo SPEAR, l'anello di accumulazione elettroni-positoni di Stanford, per ottenere la luminosità record di 10^{31} particelle per centimetro quadrato per secondo. Con tale luminosità si possono produrre reazioni con sezioni d'urto fino a 10^{-34} centimetri quadrati alla frequenza di circa un evento ogni 17 minuti.

Quali sono gli esperimenti che si possono eseguire con queste macchine? Conviene dividere gli esperimenti in quattro gruppi: 1) verifiche della validità della teoria sull'elettrodinamica quantistica; 2) ricerca di nuove particelle; 3) esperimenti con fasci di « fotoni equivalenti »; 4) produzione di particelle a interazione forte.

L'elettrodinamica quantistica è la teoria dei fenomeni elettromagnetici per la quale a Richard P. Feynman, Julian Schwinger e Sin-itiro Tomonaga fu assegnato nel 1965 il premio Nobel per la fisica. Tale teoria fonde le leggi dell'elettromagnetismo classico con la teoria speciale della relatività e con la meccanica quantistica e abbraccia fenomeni cosí diversi come le forze di natura chimica tra le molecole delle cellule viventi, la produzione di radioonde e la creazione di coppie elettrone-positone con raggi gamma di alta energia. Si tratta forse della teoria fisica piú completa e meglio verificata sia per quanto riguarda la precisione delle previsioni che la loro vastità. È ovviamente molto importante per le nostre conoscenze fondamentali della natura verificare la validità dell'elettrodinamica quantistica sia con maggiore precisione che in nuovi campi della fisica.

Gli esperimenti di collisione tra fasci di elettroni e di positoni costituiscono un modo particolarmente comodo di verificare la validità della teoria dell'elettrodinamica quantistica nel campo delle alte energie. Per queste prove sono stati studiati tre tipi di reazioni: 1) reazioni in cui gli elettroni e i positoni schizzano via nell'urto senza perdere energia (processo chiamato diffusione elastica); 2) reazioni in cui gli elettroni e i positoni si annichilano per formare coppie muone-antimuone; 3) reazioni in cui gli elettroni e i positoni si annichilano per produrre coppie di raggi gamma.

La diffusione elastica di elettroni da parte di positoni dipende tanto dalla struttura interna delle particelle quanto dalla natura delle forze agenti. Le attuali ipotesi considerano l'elettrone e il positone come cariche puntiformi (senza alcuna estensione spaziale) che possono interagire solo attraverso forze elettriche e magnetiche a lunga distanza. Tali ipotesi sono in accordo con tutti i risultati sperimentali ottenuti finora, ma vale la pena di verificarle a energie sempre piú grandi o, il che è la stessa cosa, a distanze sempre piú piccole.

Nonostante lo spettacolare successo dell'elettrodinamica quantistica vi sono ancora alcuni problemi fondamentali sull'elettrone rimasti insoluti. Per esempio, la carica dell'elettrone viene spesso espressa in funzione di una quantità adimensionale α (chiamata co-

La vista trasversale di un evento nel BOLD mostra una diffusione elastica di un positone da parte di un elettrone. Le due particelle si allontanano in verso opposto. Tale risultato è la conseguenza del principio di conservazione della quantità di moto, secondo il quale la quantità di moto totale di un sistema isolato si mantiene costante nel tempo. Sebbene si sia indicata una particella come elettrone e l'altra come positone non è possibile individuare quale sia l'una e quale l'altra.

La vista in piano di una diffusione di un positone da parte di un elettrone corrisponde allo schema della pagina a fianco, ma è perpendicolare alla direzione del fascio anziché parallela. Questa vista mostra non solo che il positone viene diffuso in verso opposto a quello dell'elettrone, ma anche che le due particelle formano un angolo con la loro direzione di moto iniziale. Questo angolo, detto angolo di diffusione, varia da 50 a 130 gradi negli esperimenti di Cambridge. Qui è di 110 gradi.

stante di struttura fine) che è numericamente uguale a circa 1/137. Questo numero misterioso è una delle costanti fondamentali della natura, tuttavia non può essere ricavato né dall'elettrodinamica quantistica né da altre teorie. Non è nemmeno chiaro perché le cariche elettriche appaiono sempre in natura come multipli interi della carica elementare dell'elettrone. Nello stesso modo la massa dell'elettrone non può essere calcolata con i metodi dell'elettrodinamica quantistica né di qualsiasi altra teoria: è una proprietà il cui valore deve essere determinato sperimentalmente. È auspicabile che qualche futura teoria dell'elettrodinamica sia in grado di spiegare l'origine della carica e della massa dell'elettrone ed è con un occhio rivolto a tali possibili sviluppi che i ricercatori stanno indagando sulla struttura dell'elettrone a distanze sempre minori.

Rivolgiamo ora la nostra attenzione allo studio sperimentale della reazione di diffusione elastica elettrone-positone. Quando un elettrone si avvicina a un positone (per esempio, all'interno di un anello di accumulazione) ognuna delle due particelle esercita sull'altra delle forze che la fanno diffondere, o deviare, dalla direzione iniziale di moto. L'angolo di diffusione di una particella è definito come l'angolo tra la direzione iniziale e quella finale.

Negli esperimenti di collisione di fasci trattiamo il caso di un fascio di elettroni e di uno di positoni della stessa energia lanciati uno contro l'altro. Viene contato il numero di coppie elettrone-positone che esce nell'unità di tempo dalla zona di interazione in un dato intervallo di angoli di diffusione. Da questo numero, noto il valore della luminosità media, si può calcolare la probabilità di diffusione, o sezione d'urto. La teoria dell'elettrodinamica quantistica predice la sezione d'urto di diffusione in funzione dell'energia e dell'angolo di diffusione; per verificare la validità di tale teoria vengono confrontate le previsioni con i risultati sperimentali.

Recentemente è stato eseguito presso il Cambridge Electron Accelerator, da un gruppo di fisici guidati da Robert Little, un esperimento di diffusione elastica elettrone-positone. (Noi siamo due dei membri di quel gruppo). Vennero fatti collidere un fascio di elettroni da 2000 MeV e uno di positoni da 2000 MeV con un'energia totale nel centro di massa di 4000 MeV. La zona di interazione venne circondata con un'apparecchiatura di rivelazione costituita da camere a scintilla, contatori a scintillazione e lastre metalliche disposte in modo da formare quattro quadranti identici (*si veda lo schema a pag. 10*). In due mesi di raccolta di dati sono stati rivelati 260 eventi che potrebbero essere attribuiti a processi di diffusione elastica, con angoli di diffusione compresi tra 50 e 130 gradi. (In un tipico evento di questo genere è impossibile distinguere l'elettrone dal positone, dato che il rivelatore è privo di campo magnetico).

Le due tracce registrate nelle camere a scintilla da un evento di questo tipo appaiono allineate (*si vedano gli schemi nelle due pagine precedenti*). Questo risultato è una conseguenza del principio di conservazione della quantità di moto, che afferma che la quantità di moto totale di un sistema isolato non può cambiare nel tempo. In questa particolare reazione di fasci in collisione la quantità di moto totale prima dell'urto è nulla, poiché la quantità di moto dell'elettrone iniziale è esattamente bilanciata dalla quantità di moto uguale e opposta del positone iniziale. La conservazione della quantità di moto richiede perciò che anche la quantità di moto totale dopo l'urto sia nulla e ciò può accadere solo se le quantità di moto delle particelle finali sono uguali e opposte. Di conseguenza l'elettrone e il positone devono muoversi in versi opposti dopo la reazione di diffusione.

Le particelle finali possono produrre sulle lastre metalliche degli sciami elettromagnetici; tali sciami vengono prodotti secondo il meccanismo che abbiamo descritto a proposito della produzione di positoni. La presenza di tali sciami nelle registrazioni della camera a scintilla permette di identificare inequivocabilmente la reazione come una diffusione elastica di elettroni e positoni.

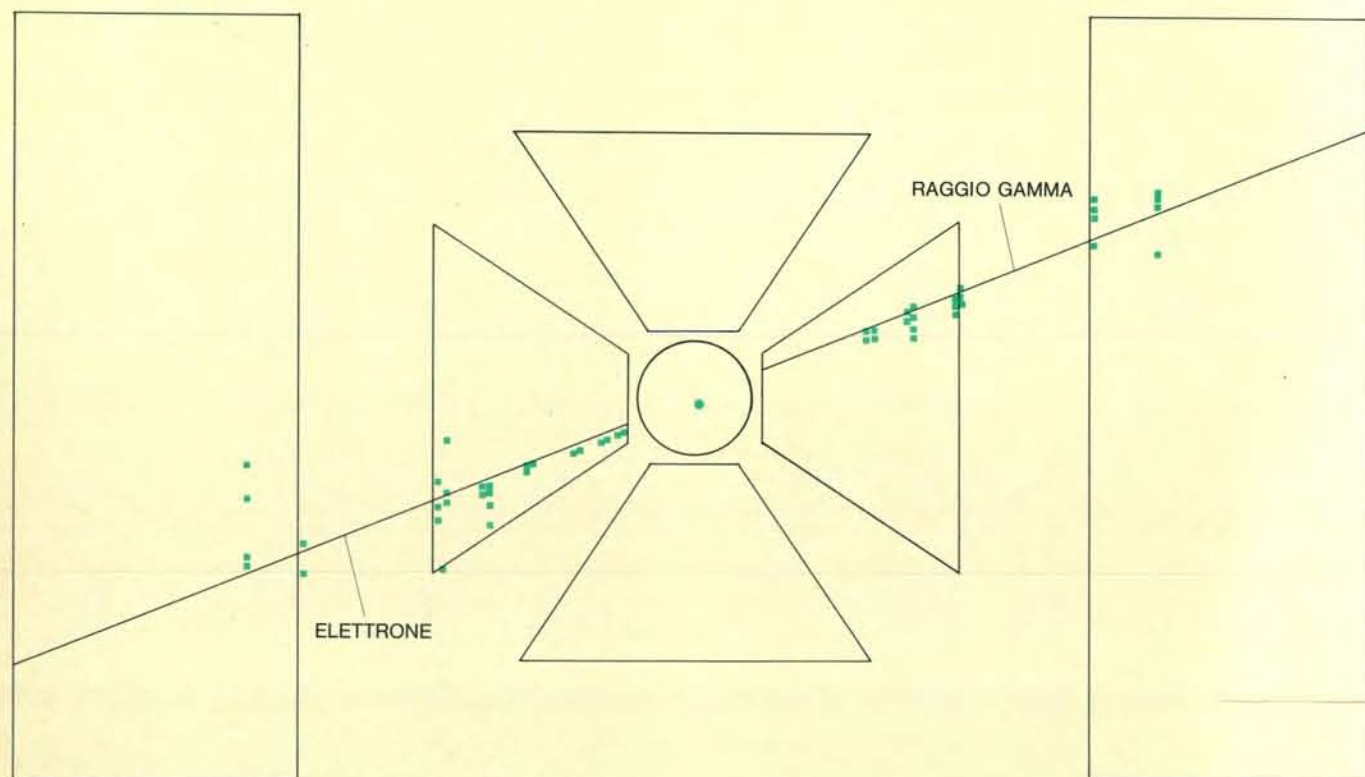
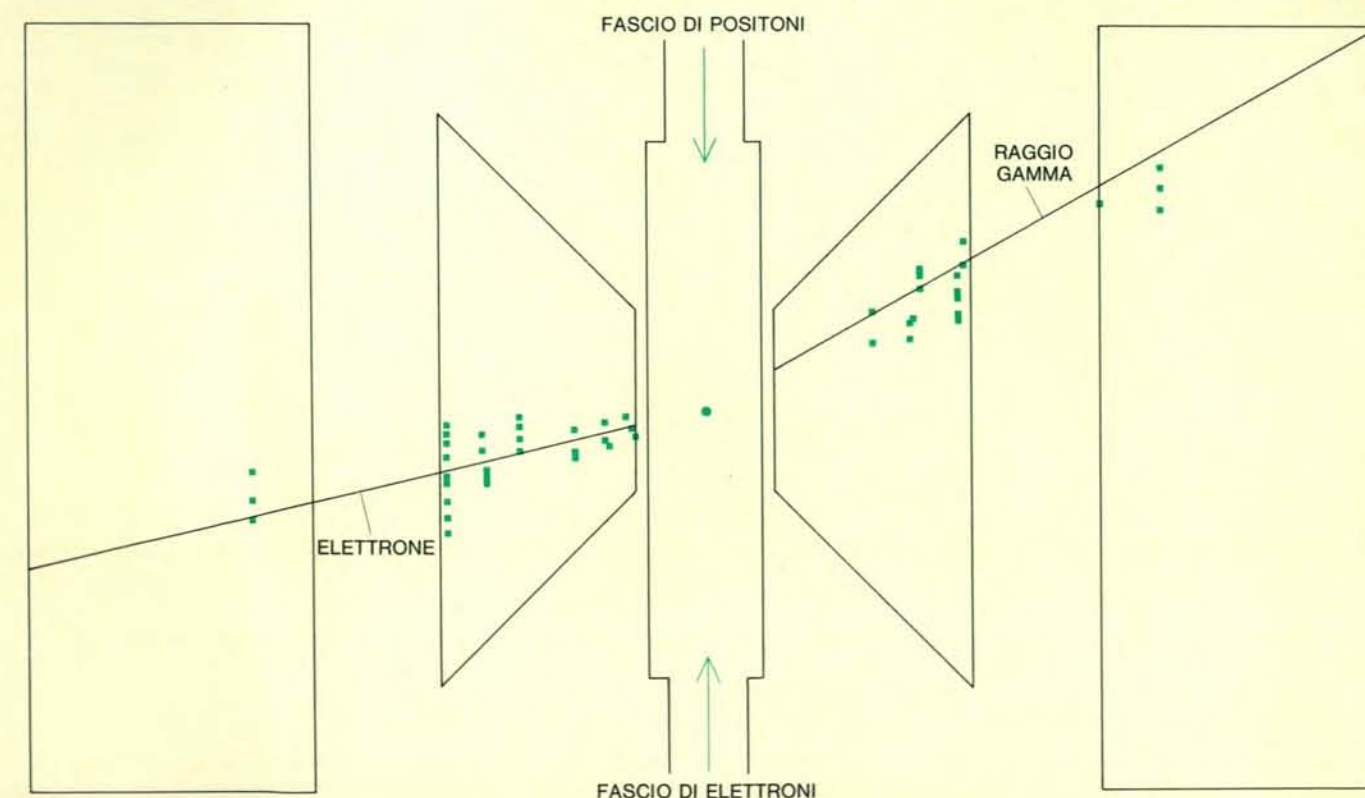
Prima di presentare i risultati dei nostri esperimenti cercheremo di spiegare i punti principali delle previsioni teoriche. Nell'elettrodinamica classica la forza responsabile della diffusione di una particella carica da parte di un'altra è la forza elettrica della legge di Coulomb. Nell'elettrodinamica quantistica tale forza è espressa in termini di uno scambio di fotoni tra le particelle interagenti.

Il fotone è il quanto della radiazione elettromagnetica. Un fascio di luce o di radioonde è costituito da molti fotoni; un raggio gamma è un singolo fotone di alta energia. I fotoni vengono emessi tutte le volte che una particella carica viene accelerata o decelerata. Tutte le interazioni tra elettroni e fotoni – per esempio l'emissione (o l'assorbimento) di luce da un atomo o la produzione (o ricezione) di radioonde da parte degli elettroni di un'antenna – sono basate su una singola reazione del tipo chiamato reazione « virtuale ». La rappresentazione convenzionale che i fisici danno di tale reazione è detta diagramma di Feynman (*si veda lo schema a pag. 11*).

Una reazione virtuale non può mai essere osservata in pratica, dato che in essa non si possono conservare contemporaneamente l'energia e la quantità di moto. È tuttavia possibile provocare una successione di reazioni virtuali in modo da produrre processi reali, osservabili (nei quali devono naturalmente conservarsi sia l'energia sia la quantità di moto). Supponiamo, per esempio, che il primo elettrone emetta un fotone. Il fotone ha massa nulla, quindi tale reazione può conservare la quantità di moto, ma non può conservare l'energia; secondo il principio di indeterminazione questo duplice risultato è possibile solo se il fotone esiste per un tempo brevissimo. Alternativamente (in modo equivalente) si può supporre che vengano conservate sia l'energia che la quantità di moto attribuendo al fotone virtuale una massa fittizia. L'assorbimento del fotone da parte di un secondo elettrone è un altro esempio di processo virtuale. Il processo complessivo può essere interpretato come l'interazione di un elettrone con un altro elettrone mediante forze elettromagnetiche; in tal caso nella reazione si conservano sia l'energia che la quantità di moto.

Il fotone scambiato tra il primo e il secondo elettrone non compare tra le particelle iniziali o finali e quindi non potrà mai essere osservato direttamente (e questo è il motivo per cui è chiamato virtuale). Tale fotone virtuale scambiato può però trasportare energia e quantità di moto. La diffusione dei due elettroni viene descritta come uno scambio di un fotone virtuale con trasferimento di quantità di moto dall'una all'altra particella.

Anche il positone rientra in questo schema. Feynman ha dimostrato che l'emissione virtuale di un fotone da parte di un positone segue le stesse modalità del processo descritto per l'elettrone. La sua analisi ha portato alla predizione di un processo ben al di fuori del dominio della fisica pre-quantomeccanica: l'annichilazione virtuale di una coppia elettrone-positone per formare un fotone.

Si possono combinare questi processi virtuali per ricavare la rappresentazione secondo l'elettrodinamica quantistica della diffusione di un elettrone da parte di un positone mediante i due diagrammi di Feynman alternativi. In un diagramma l'elettrone emette un fotone virtuale che viene assorbito dal positone. Nell'altro diagramma l'elettrone e il positone si annichilano per formare un fotone virtuale che viene trasformato in una coppia elettrone-positone. Assegnando fattori numerici

L'evento, chiamato diffusione Compton virtuale, ha un elettrone e un raggio gamma nello stato finale (*a sinistra*). Un fascio di positoni nel processo di collisione elastica con un fascio di elettroni può essere trattato come un fascio di « fotoni equivalenti ». Nella diffusione virtuale Compton il positone che ha prodotto il fotone equivalente subisce una degradazione energetica e si muove prevalentemente in avanti. Nella vista perpendicolare ai fasci in collisione (*a destra*) l'elettrone non si muove in verso opposto a quello del raggio gamma, a causa della quantità di moto portata via dal positone che si muove in avanti. La reazione avrebbe potuto aver luogo ugualmente se il fotone equivalente avesse fatto diffondere un positone e non un elettrone.

a ogni ramo, a ogni vertice e a ogni fotone virtuale del diagramma si possono calcolare le previsioni teoriche per la sezione d'urto di reazione.

Esaminiamo ora i risultati sperimentali ottenuti dal nostro gruppo di Cambridge. Il numero totale di eventi di diffusione elastica rivelati e la distribuzione degli angoli di diffusione delle particelle finali sono in accordo con le previsioni dell'elettrodinamica quantistica. Tale accordo è stato confermato anche in esperimenti eseguiti a energie inferiori.

Un'altra particella interessante prodotta nella collisione di fasci di elettroni e positoni è il muone. Questa particella ha la stessa carica e lo stesso spin dell'elettrone, ma è circa 200 volte piú pesante dell'elettrone. Un muone libero decade con una vita media di due microsecondi in un elettrone, un neutrino e un antineutrino. I muoni, come gli elettroni, non interagiscono con i nuclei con interazioni nucleari forti. Come l'elettrone, anche il muone appare come una carica puntiforme. Se l'energia è abbastanza alta, i muoni possono essere prodotti a coppie mediante reazione di annichilazione di un elettrone e di un positone. Questa reazione procede anch'essa mediante scambio di un fotone virtuale e può essere rappresentata con un diagramma di Feynman.

Le coppie muone-antimuone possono essere facilmente identificate negli esperimenti. Nell'attraversare la materia i muoni irradiano meno degli elettroni; essi non interagiscono fortemente con i nuclei e possono quindi percorrere distanze notevoli. Poiché il muone è facilmente identificabile, una coppia è ancora piú facilmente identificabile. Coppie muone-antimuone sono state prodotte in tutti i laboratori di anelli di accumulazione e i risultati ottenuti sono in accordo con l'elettrodinamica quantistica. Particelle come i muoni, gli elettroni e i neutrini, che hanno tutte spin 1/2 e nessuna delle quali subisce l'intensa forza nucleare, sono chiamate leptoni.

Un'altra verifica dell'elettrodinamica quantistica si può ottenere esaminando la reazione nella quale un elettrone e un positone si annichilano producendo una coppia di raggi gamma. In questo processo viene scambiato un elettrone virtuale. Anche i raggi gamma vengono facilmente identificati; essi producono sciami elettromagnetici nel piombo, ma non vi sono tracce di particelle cariche che escono dalla zona di interazione. Ancora una volta le previsioni teoriche sono in accordo con i risultati sperimentali.

Le collisioni di fasci di elettroni e positoni costituiscono un valido strumento nella ricerca di nuove particelle capaci di interagire con elettroni o fotoni. Le previsioni indicano un grande numero di siffatte particelle; la dimostrazione della loro esistenza dovrebbe essere utile nella formulazione di modelli teorici delle interazioni dei leptoni e delle particelle elementari. Molte di esse potrebbero essere prodotte in linea di principio in collisioni elettrone-positone di energia sufficientemente alta e potrebbero essere facilmente identificate se esistono realmente. Ne sono ipotetici esempi un elettrone eccitato o un leptone pesante.

Atomi, nuclei e nucleoni (protoni e neutroni) possono essere eccitati dai loro stati fondamentali, o di minima energia, a stati di energia maggiore mediante assorbimento di un fotone. Non si sa se anche elettroni e muoni hanno stati eccitati. Se tali stati dovessero esistere, si potrebbero produrre con collisioni elettrone-positone e potrebbero essere rivelati mediante i loro decadimenti caratteristici. Se si dovessero scoprire tali stati eccitati, essi potrebbero indicare che l'elettrone e il muone non sono entità « elementari » ma posseggono una struttura complessa.

L'elettrone e il muone (con le loro antiparticelle) sono i soli leptoni noti elettricamente carichi. Il motivo per cui vi dovrebbe essere piú di uno ma soltanto due leptoni carichi è uno degli interrogativi piú vecchi della fisica delle particelle. Una possibilità è che vi siano piú di due leptoni carichi, ma che solo l'elettrone e il leptone siano stati scoperti finora. In effetti potrebbe anche esistere uno spettro di leptoni di massa crescente. I leptoni pesanti, se esistono, si potrebbero produrre, come i muoni, in coppie nelle reazioni di annichilazione elettrone-positone. Essi dovrebbero decadere in un elettrone (o un muone), un neutrino e un antineutrino, o in un neutrino e una o piú particelle a interazione forte. I modi di decadimento dovrebbero identificare la particella. La scoperta di un nuovo leptone potrebbe chiarire il motivo della differenza di massa tra l'elettrone e il muone o, piú in generale, le relazioni esistenti tra l'elettrone e il muone. Attualmente non vi è alcuna evidenza dell'esistenza di nessuna di tali particelle ipotetiche.

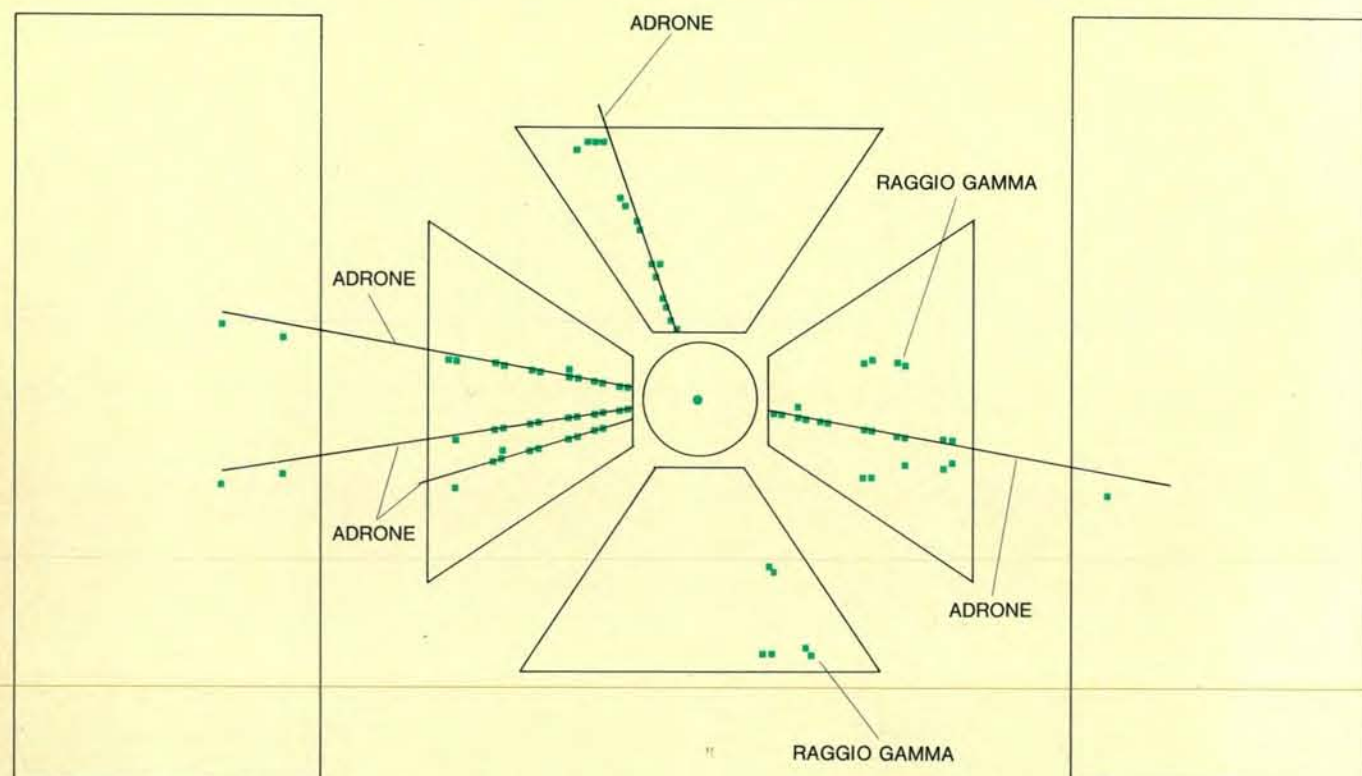
Collisioni tra fasci elettrone-elettrone o elettrone-positone possono anche essere utilizzate per studiare collisioni fotone-fotone. Questa possibilità è stata sottolineata negli ultimi anni dai teorici sovietici, francesi e statunitensi. La teoria di questi processi è basata sull'equivalenza approssimata tra i campi elettrici e magnetici prodotti da una particella carica in rapido movimento e i campi prodotti da un impulso di radiazione elettromagnetica. Ognuno di questi impulsi può essere decomposto (con l'analisi di Fourier) in uno spettro di fotoni. L'interazione di una particella carica in moto con la materia può allora essere calcolata approssimativamente in funzione dell'interazione della materia dello spettro equivalente di fotoni. Questa « approssimazione di fotoni equivalenti » venne formulata indipendentemente da C.F. von Weizsäcker e da E.J. Williams nel 1934. Secondo tale metodo un fascio di positoni in collisione con un fascio di elettroni può essere trattato come un fascio di fotoni equivalenti. Uno dei fotoni equivalenti può far diffondere elasticamente un elettrone dando origine a un elettrone e a un raggio gamma nello stato finale; nello stesso tempo il positone che ha prodotto il fotone equivalente viene degradato energicamente e si muove prevalentemente in avanti.

Eventi di questo tipo sono stati registrati nelle camere a scintilla (*si vedano gli schemi nelle pagine 14 e 15*). Nella vista perpendicolare l'elettrone e il raggio gamma non si muovono in versi opposti sulla stessa direzione, a causa della quantità di moto trasportata dal positone che viaggia in avanti. Sedici di tali eventi sono stati registrati a Orsay e sei a Cambridge con energie nel centro di massa rispettivamente di 700 e di 3000 MeV; i risultati sono in accordo con le previsioni dell'elettrodinamica quantistica.

Entrambi i fasci di particelle cariche possono essere trattati come fasci di fotoni equivalenti: questo fatto suggerisce la possibilità di osservare reazioni nelle quali fotoni equivalenti entrano in collisione per produrre coppie elettrone-positone, muone-antimuone o particella-antiparticella a interazione forte. Le sezioni d'urto delle prime due reazioni si possono calcolare con le tecniche dell'elettrodinamica quantistica. Eventi di questo tipo sono stati osservati a Novosibirsk, a Frascati e a Cambridge e i risultati sono in accordo con i calcoli teorici. La terza reazione offre interessanti prospettive per futuri esperimenti.

Attualmente il campo piú interessante di ricerca con collisioni di fasci elettrone-positone è la produzione di particelle nucleari a interazione forte chiamate adroni. Queste particelle, a causa delle loro interazioni forti, hanno struttura e dimensioni finite e le loro interazioni elettromagnetiche non possono essere descritte con una semplice applicazione dell'elettrodinamica quantistica.

Gli adroni possono essere prodotti in coppie, come nell'annichilazione di un elettrone e di un positone per produrre un protone e un antiprotone. La probabilità di produzione di una coppia di adroni in una collisione elettrone-positone viene espressa mediante un gruppo di funzioni chiamate fattori di forma, che risultano funzioni della quantità di moto trasferita al fotone virtuale. I fattori di forma che descrivono la produzione di una coppia di adroni, con una lieve estrapolazione, dovrebbero anche descrivere la diffusione elastica di elettroni da parte dell'adrone. Il fattore di forma è quindi una misura della distribuzione spaziale della carica, o del momento magnetico, dell'adrone.

L'adrone studiato piú a fondo nelle collisioni elettrone-positone è il pione; esso è la piú leggera delle particelle a interazione forte e ha una massa che è solo 280 volte quella dell'elettrone. Se il pione fosse una particella puntiforme il quadrato del suo fattore di forma dovrebbe essere uguale a 1; deviazioni dall'unità sono causate dalle sue interazioni forti (*si veda lo schema in basso a pagina 18*). L'aspetto piú sorprendente del grafico del fattore di forma del pione è il picco a un'energia nel centro di massa di 760 MeV con una larghezza di 120 MeV. Esso

Gli adroni (particelle nucleari a interazione forte) vengono prodotti in collisioni tra elettroni e positoni alle alte energie. Uno degli eventi osservati al BOLD presentava cinque particelle cariche, probabilmente adroni, nello stato finale.

viene interpretato come uno stato risonante formato dalla coppia pione-antipione. La risonanza può essere considerata come una particella a breve vita media, chiamata mesone ro neutro, che decade in 10^{-23} secondi in una coppia pione-antipione. Cosí, in corrispondenza a energie nel centro di massa corrispondenti all'incirca alla massa del mesone ro neutro (760 MeV) sulla produzione di coppie di pioni predomina la reazione di annichilazione elettrone-positone con produzione di un mesone ro neutro, che decade a sua volta producendo una coppia pione-antipione.

Nelle collisioni elettrone-positone, oltre al mesone ro neutro sono state prodotte altre due risonanze ben identificate, il mesone omega neutro e il mesone fi neutro; gli esperimenti di collisione di fasci si sono rivelati il modo piú semplice di studiare queste tre risonanze.

Il diagramma di Feynman per i quark, entità teoriche che potrebbero essere l'ultimo stadio di particelle subatomiche, mostra come tali entità potrebbero spiegare la formazione degli adroni nelle collisioni tra fasci di positoni e di elettroni. Un positone e un elettrone dovrebbero annichilarsi fino a formare un quark e un antiquark.

Il pione non è una particella puntiforme, come si vede in un grafico che rappresenta il quadrato del fattore di forma del pione in funzione dell'energia nel centro di massa del sistema formato dal pione e dalla sua antiparticella. Il fattore di forma è una misura della distribuzione spaziale della carica di un adrone. Se il pione fosse una particella puntiforme, il quadrato del suo fattore di forma dovrebbe essere uguale a 1. Le deviazioni dall'unità sono dovute alle sue interazioni forti. Il grande picco a 760 MeV è interpretato come uno stato di risonanza formato tra il pione e la sua antiparticella. I dati riportati sono tratti dagli esperimenti eseguiti a Novosibirsk in URSS.

Ha destato molto interesse la constatazione, prima a Frascati e poi a Cambridge, che gli adroni vengono prodotti nelle collisioni elettrone-positone ad alta energia molto piú spesso di quanto si prevedesse. In questi esperimenti si sono osservati eventi di produzione di molte particelle cariche (*si veda lo schema a fronte*). Si sta intraprendendo un notevole sforzo per indagare a fondo sulla natura di questi eventi multiadronici.

Per energie nel centro di massa inferiori a 1100 MeV le risonanze ro, omega e fi predominano sulla produzione di adroni. L'estrapolazione a energie piú alte era incerta prima dei recenti esperimenti. L'opinione prevalente era che il rapporto di produzione tra adroni e coppie muone-antimuone fosse di circa 1: 1; le misure eseguite a Frascati e a Cambridge mostrano che tale rapporto è di circa 1: 1 per un'energia nel centro di massa di 1350 MeV per crescere fino a piú di 4: 1 a 4000 MeV. Ciò costituisce una grossa sorpresa e dimostra che la fisica che sta dietro a una grande sezione d'urto di produzione di adroni è tuttora sconosciuta.

Una possibile spiegazione è che altre risonanze siano responsabili della grande sezione d'urto di produzione degli adroni. A Frascati vi sono evidenti tracce di una nuova risonanza, chiamata ro primo, in corrispondenza a una massa di circa 1600 MeV; tale risonanza decade in quattro pioni ed è stata osservata anche in esperimenti di produzione di fotoni. Il modello a « quark » delle particelle elementari proposto qualche anno fa da Murray Gell-Mann induce a prevedere altre risonanze come « compagni » del ro primo che facciano parte dello stesso « nonetto ». Tali risonanze dovrebbero anche contribuire alla produzione di adroni. È probabile che esistano altre risonanze ancor piú pesanti che si sovrappongono in uno spettro continuo producendo la elevata sezione d'urto.

Un'altra interpretazione che fornisce una previsione con la quale confrontare i nostri risultati è il modello a « partoni » di Feynman. Esperimenti di diffusione anelastica elettrone-protone eseguiti a Stanford paiono evidenziare un tipo di diffusione da parte di tre costituenti puntiformi del protone chiamati partoni; tali entità, per rendere conto dei risultati, devono avere spin 1/2; i risultati consentono grosso modo di identificare i partoni con

	PRIMA DELLA COLLISIONE		DOPO LA COLLISIONE
	PRIMA AUTO	SECONDA AUTO	SCONTRO
	IN MOTO VERSO DESTRA	FERMA	IN MOTO VERSO DESTRA A VELOCITÀ DIMEZZATA
(BERSAGLIO FISSO)			
QUANTITÀ DI MOTO	$+ mv$	0	$(2 m) (v/2) = mv$
ENERGIA CINETICA	$mv^2/2$	0	$1/2 (2 m) (v/2)^2 = mv^2/4$

L'urto contro un bersaglio fisso simula la situazione che si ha in un acceleratore di particelle di tipo tradizionale. Entrambe le auto, rappresentate nella figura, hanno massa m e una soltanto di esse si muove a velocità v. Quando avviene l'urto solo un quarto dell'energia originaria dell'auto in movimento ($mv^2/4$) viene utilizzata per schiantare l'altra auto.

	IN MOTO VERSO DESTRA	IN MOTO VERSO SINISTRA	URTO E ARRESTO
(COLLISIONE)			
QUANTITÀ DI MOTO	$+ mv$	$- mv$	0
ENERGIA CINETICA	$mv^2/2$	$mv^2/2$	0

L'urto tra due oggetti in movimento simula la situazione che si realizza sfruttando il principio degli anelli di accumulazione. L'energia complessiva delle due automobili, mv^2, viene completamente utilizzata per schiantare il metallo. Si tratta di una energia che è ben quattro volte maggiore di quella ottenibile nel caso del bersaglio fisso; da ciò si comprende come il perfezionamento tecnologico che consente di disporre di energie di fasci sempre maggiori sarebbe stato sfruttabile solo in minima parte, per l'esistenza di invalicabili limiti fisici, se non vi fosse stata l'invenzione degli anelli di accumulazione. I due fasci si muovono uno incontro all'altro con velocità del centro di massa nulla, quindi, anche dopo l'urto, tale velocità sarà nulla e nulla sarà di conseguenza l'energia del centro di massa: tutta l'energia dei due fasci sarà utilizzabile nella reazione nucleare.

quark di carica frazionaria di $-1/3$, $-1/3$ e $2/3$. Il successo con cui il modello a partoni descrive questi esperimenti ha destato molto interesse nel verificare se questo modello si applica anche alle reazioni di annichilazione.

Nel modello a partoni il processo di produzione degli adroni può essere schematizzato nel modo seguente: un elettrone e un positone si annichilano producendo un partone e un antipartone, che vengono poi trasformati in adroni. Il diagramma di Feynman per questo processo è simile al diagramma di annichilazione di elettroni e positoni in muoni.

La sezione d'urto di produzione di adroni può quindi essere calcolata proprio come la sezione d'urto di produzione muone-antimuone (almeno se trascuriamo gli effetti delle interazioni negli stati finali). Il muone e il partone sono entrambi puntiformi e hanno spin 1/2; essi possono però differire per la carica elettrica. Possiamo perciò ottenere una previsione quantitativa: il rapporto tra la produzione di adroni e la produzione di coppie di muoni alle alte energie è uguale alla somma dei quadrati delle cariche dei partoni; nel normale modello a quark tale rapporto risulta di 2:3. Talvolta si suppone l'esistenza di un gruppo piú numeroso di quark, chiamati quark colorati e in tal caso il rapporto diventa di 2:1. I risultati sperimentali sono in disaccordo con questi modelli.

Un'altra previsione del modello a partoni è che gli adroni dovrebbero uscire dalla reazione in coni all'indietro chiamati « jet »; per ora non vi è alcuna prova di strutture a jet negli eventi degli adroni, ma alle energie attualmente disponibili tali strutture non pare possano essere facilmente messe in evidenza. Se i jet esistono, dovrebbe essere piú facile rivelarli a energie maggiori.

Si vede, perciò, che gli esperimenti in corso ci hanno solo stimolati risvegliando il nostro appetito per ulteriori esperimenti. Esistono i leptoni pesanti? Vi sono altre risonanze? Si scopriranno le strutture a jet? Il modello a partoni è valido a qualsiasi energia? Cosa c'è dietro alla grande sezione d'urto di produzione degli adroni? Le risposte a questi interrogativi sono importanti ai fini delle nostre conoscenze sulle particelle elementari. I fisici attendono i risultati dei prossimi esperimenti di collisione di fasci di elettroni e positoni.

Virus erpetici e cancro

È noto da tempo che questi virus ubiquitari possono provocare il cancro negli animali da esperimento; ora è stato dimostrato che virus di questo tipo sono implicati anche in alcuni tipi di cancro dell'uomo

di Keen A. Rafferty jr.

I virus erpetici sono tra i virus piú diffusi trovati nell'uomo. Essi hanno la misteriosa capacità di comparire e scomparire negli individui senza che ne sia conosciuta la via di infezione o la causa. Uno di questi virus, l'*Herpes simplex* Tipo 1, produce le caratteristiche vescicole da « febbri » della bocca e colpisce praticamente tutta la popolazione umana. L'*Herpes simplex* Tipo 2 è invece trasmesso attraverso i rapporti sessuali e provoca piaghe sui genitali. Un altro virus erpetico comune è l'*Herpes zoster*, che provoca la varicella nei bambini e ricompare sotto forma di fuoco di Sant'Antonio negli adulti ed anche nei bambini. Tutti questi virus appartengono a un gruppo comprendente forme capaci di provocare il cancro in animali da esperimento. Sebbene i virus dell'*Herpes simplex* e dell'*Herpes zoster* sembra non provochino il cancro nell'uomo, ci sono buone – sebbene ancora non determinanti – prove secondo le quali alcuni tipi di cancro estremamente maligni dell'uomo siano causati, o prevalentemente associati a virus del gruppo erpetico. Poiché i virus erpetici possono essere talvolta isolati da infezioni locali o da cellule tumorali e fatti crescere in colture di tessuto, possono rappresentare il mezzo migliore per provare il ruolo eziologico dei virus nel cancro umano, offrendo inoltre la possibilità di preparare vaccini per prevenire alcune forme di cancro, proprio come ora si previene la poliomelite.

La parola « virus » deriva dal latino e vuol dire veleno. Nel suo significato moderno fu usata per la prima volta da Louis Pasteur che la applicò a quegli agenti patogeni che non erano batteri pur essendo associati a malattie. Sebbene si sia subito riconosciuto che molte malattie potevano essere provocate da virus, è stata invece accettata con difficoltà l'ipotesi che essi potessero causare anche il cancro. Non è facile capire questo fatto tenendo presenti le prove inequivocabili che si sono ottenute da tempo grazie agli esperimenti su animali. Nel 1908 V. Ellerman e O. Bang dimostrarono che la leucemia dei polli poteva essere trasmessa iniettando dei filtrati privi di batteri provenienti da polli infetti in polli sani. Tre anni dopo Peyton Rous dimostrò che la stessa cosa si verificava per il sarcoma dei polli, un tumore maligno dei tessuti profondi. Rous ricevette per questo lavoro il premio Nobel, ma è significativo il fatto che il riconoscimento gli fu conferito 55 anni dopo l'importante scoperta.

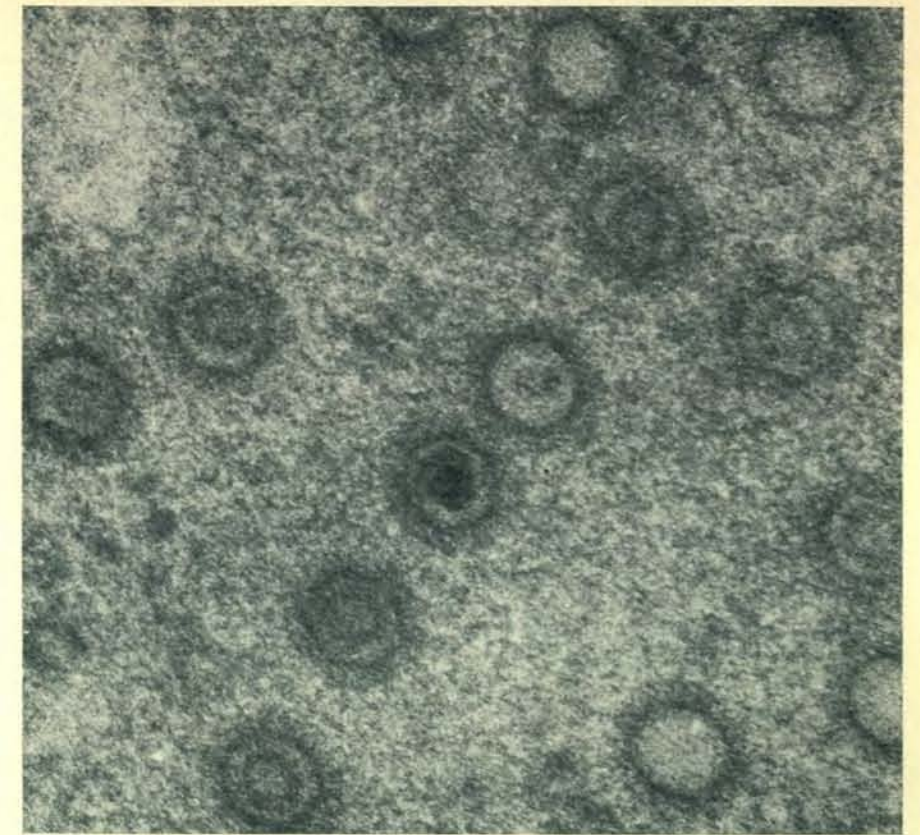
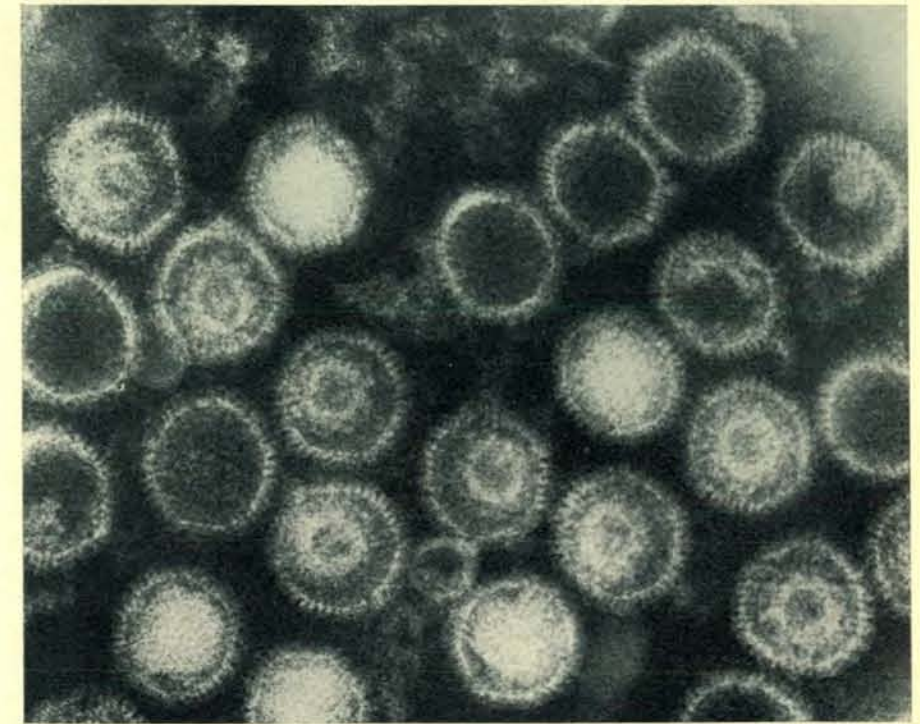
Una ragione per essere riluttanti nell'accettare l'ipotesi che il cancro può essere causato da virus è che il cancro non è contagioso, mentre altre infezioni virali come il comune raffreddore e l'influenza lo sono. Inoltre, alcune forme di cancro dell'uomo sono, senza possibilità d'errore, associate ad agenti che non sono virus. Per esempio, è noto da 200 anni che negli spazzacamini in Inghilterra era frequente il carcinoma dei testicoli e che il catrame contenuto nella fuliggine dei camini ne era la causa probabile. Nel 1918, K. Yamagiwa e K. Ichikawa in Giappone riuscirono a far sviluppare il cancro spennellando con derivati di catrame minerale le orecchie di conigli, dimostrando cosí che il cancro può essere prodotto da sostanze chimiche. Si sono rivelate cause di cancro anche i raggi X e le radiazioni di sostanze radioattive. Nelle operaie che dipingevano con il radio i quadranti degli orologi si sviluppava il cancro del labbro perché esse appuntivano le estremità dei pennelli mettendole in bocca. Nei radiologi che lavoravano con i raggi X, c'era una tendenza a sviluppare il cancro della pelle e altre forme di cancro prima che prendessero piede delle severe norme di sicurezza. Piú recentemente, la resistenza nell'accettare la teoria virale del cancro è derivata dalla difficoltà di dimostrare la presenza di virus nei tumori, particolarmente in quelli dell'uomo.

Ci sono due tipi principali di virus: quelli che contengono acido desossiribonucleico (DNA) e quelli che contengono acido ribonucleico (RNA). Nei tumori animali causati da virus ad RNA, il virus può essere riprodotto in continuazione e nuovi virus possono essere liberati senza danneggiare le cellule ospiti. Si crede che i virus ad RNA siano i fattori responsabili di alcune forme di leucemia nell'uomo, nelle quali i globuli bianchi o rossi si dividono con rapidità prodigiosa. Tuttavia con i virus a DNA, la replicazione del virus e la sua liberazione dalla cellula ospite sembrano sempre essere accompagnate dalla morte della cellula. Quindi se un virus a DNA si dovesse replicare in un tumore, ci sarebbe la distruzione delle cellule tumorali e alla fine dovrebbe scomparire anche il virus stesso.

I virus erpetici sono virus a DNA. Sono noti due altri tipi di virus a DNA che provocano tumori. Si tratta degli adenovirus che sono i fattori responsabili del raffreddore e di altre infezioni respiratorie e dei virus papova (papilloma-polioma-vacuolante) che possono essere i fattori responsabili di una varietà di tumori (comprese le verruche, escrescenze benigne che possono essere trasmesse da persona a persona mediante iniezione del virus).

Il fatto che i virus erpetici potessero essere implicati nella formazione di tumori è emerso dallo studio della curiosa epidemiologia di certi tumori nelle rane e nei polli. La specie nordamericana rana leopardo (*Rana pipiens*) è particolarmente esposta al carcinoma renale, e questo pare sia causato da speciali fattori. Nel 1932 Baldwin Lucké della Scuola di medicina dell'Università di Pennsylvania cominciò a studiare questi tumori e subito scoprí che il nucleo di alcune cellule tumorali conteneva grosse inclusioni di colore rosa pallido. Queste inclusioni potevano essere facilmente osservate al microscopio. Si credette allora (e ancora oggi si crede) che le inclusioni di quel tipo (*Cowdry* Tipo A) fossero causate da virus. La scoperta di Lucké indicava perciò chiaramente che il tumore renale della rana, ora noto come tumore di Lucké, era il risultato di un'attività virale. Sfortunatamente la comparsa di quelle caratteristiche inclusioni era sporadica e non c'era alcun modo attendibile per rilevare direttamente il virus stesso.

Nel 1956, Don W. Fawcett, che allora lavorava al Collegio di medicina della Cornell University, osservò le inclusioni tumorali di Lucké al microscopio elettronico e trovò che esse contenevano, senza possibilità di errore, delle particelle virali. In quel medesimo periodo, cominciai anch'io a cercare inclusioni simili nei tumori della rana. Per molti anni, sebbene avessi esaminato centinaia di rane allevate in laboratorio, non riuscii a trovare una sola inclusione. Infine, una osservazione fortuita di Maria Roberts, una nostra studentessa, suggerí che i tumori delle rane presentassero le inclusioni nel nucleo della cellula durante l'inverno, ma non in estate. Essa trovò inclusioni in diverse rane, che erano arrivate in spedizioni consecutive durante l'inverno e che erano state conservate per un mese in una camera fredda prima di essere esaminate. D'altro canto, tutti i tumori che avevo esaminato si erano sviluppati spontaneamente nelle rane che erano rimaste in un laboratorio caldo per diversi mesi.

Si fece quindi l'esperimento piú ovvio. Furono eseguite le biopsie dei reni delle rane di laboratorio con tumori spontanei. Queste confermarono che non era presente alcuna inclusione intranucleare. Le rane furono poi messe in un frigorifero per diversi periodi di tempo, dopo di che venivano uccise e si esaminavano i tumori del rene. Le inclusioni erano presenti nelle cellule tumorali, cosicché fummo in grado di concludere che la produzione di virus era probabilmente iniziata in seguito alla prolungata permanenza a bassa temperatura. Ulteriori esperimenti indicavano che le cellule che producevano i virus morivano con la liberazione del virus e che il virus compa-

In questa fotografia al microscopio elettronico di una sezione sottile di linfoblasta umano fornita da M.A. Epstein della Scuola di medicina dell'Università di Bristol, le particelle del virus di Epstein-Barr sono ingrandite circa 160 000 volte. Il loro aspetto è molto simile a quello delle tipiche particelle del virus erpetico. Esse hanno un profilo esagonale e sono vuote, oppure contengono una specie di nucleo a forma di anello oppure denso. Queste particelle virali sono state isolate da una coltura di tessuto di cellule tumorali provenienti da un paziente affetto da linfoma di Burkitt.

Queste particelle virali, praticamente identiche alle particelle di Epstein-Barr, sono state isolate da una coltura di tessuto di cellule tumorali di carcinoma nasofaringeo umano. Le particelle sono ingrandite circa 160 000 volte in questa fotografia di Guy de-Thé del gruppo di ricerca sul cancro della organizzazione mondiale della sanità.

Viene qui raffigurato il ciclo vitale del virus del tumore di Lucké. Nelle rane leopardo adulte infettate con il virus si sviluppano spesso carcinomi renali. Si pensa che il processo cominci quando il virus, che si trova nell'acqua dove vengono deposte le uova, infetta le uova o i girini. Nelle rane giovani, nel primo e nel secondo anno di vita, non c'è alcun segno di crescita tumorale. Verso la terza o quarta estate in alcune rane adulte si sviluppano tumori renali. Questi tumori estivi non producono virus. Le rane nelle quali si sviluppano i tumori possono morire durante l'estate o possono vivere fino all'ibernazione invernale secondo i diversi fattori ambientali. Durante l'ibernazione la crescita del tumore cessa e nel nucleo delle cellule tumorali compaiono delle inclusioni. Molte delle cellule tumorali muoiono, e questo sta a indicare che probabilmente il virus è stato attivato dalla bassa temperatura e si sta replicando. Il virus si accumula nell'urina e in primavera la rana, affetta da tumore, libera notevoli quantità di virus nell'acqua dove vengono deposte le uova, insieme all'urina. Quando fa caldo le inclusioni nelle cellule tumorali scompaiono, la crescita del tumore riprende e la rana muore. Sembra quindi chiaramente dimostrato che il tumore di Lucké viene trasmesso esclusivamente dai virus che si producono nel rene di rane affette dal tumore che vanno in ibernazione e poi sopravvivono per un inverno dopo la iniziale comparsa del tumore.

riva nell'urina, un fatto non inatteso considerando la localizzazione del tumore (*si veda l'illustrazione in queste due pagine*).

In seguito riuscimmo a isolare un virus dai tumori renali facendolo crescere in colture di tessuto di cellule di rana. Le osservazioni al microscopio elettronico mostrarono che si trattava di un virus erpetico. Tuttavia, il virus in coltura di tessuto non produceva tumori quando veniva iniettato nei girini. Questo era in contrasto con quanto era stato trovato da K.S. Tweedell dell'Università di Notre Dame, secondo il quale gli omogenati di tumori con inclusioni producevano tumori. In seguito, Maneth Gravell del St. Jude Children's Research Hospital di Memphis dimostrò che il virus in coltura di tessuto e i virus prevalenti nel tessuto tumorale avevano DNA diversi, e che sebbene essi fossero entrambi virus erpetici, non erano derivati uno dall'altro. A questo punto, si può soltanto ipotizzare che nel tumore siano presenti due virus, che il virus venga profondamente alterato nella coltura di tessuto (un evento alquanto improbabile) o che il virus della coltura provenga dalle cellule originali della coltura tissutale. È anche possibile che la quantità di virus della coltura iniettata nei girini non fosse sufficiente a provocare i tumori negli animali.

Recentemente, William Collard e i suoi collaboratori della Scuola di medicina dell'Università di Saint Louis e dell'Università di Tulane hanno mostrato che l'RNA dei tumori delle rane in fase estiva e il DNA delle inclusioni virali dei tumori della fase invernale sono in grado di formare « ibridi » vale a dire che, quando frammenti delle molecole filamentose dei due acidi nucleici si mescolano, alcuni di essi aderiscono ad altri. Questo indica che alcune parti delle due molecole sono in stretta relazione reciproca. Dato che l'RNA può essersi formato solo come copia del DNA virale, questi risultati forniscono altre prove secondo le quali nel tumore apparentemente privo di cellule è presente del materiale genetico del virus erpetico. In ogni modo ci sono pochi dubbi che il tumore di Lucké sia causato da un virus erpetico, sebbene non sia ancora certo quale virus erpetico sia implicato. Abbastanza recentemente Robert F. Naegele e Allan Granoff del St. Jude Hospital hanno completato il ciclo isolando l'agente infettivo da un tumore che era stato indotto mediante iniezione di un estratto tumorale, e usando questo allo scopo di indurre un altro tumore. Cosí, il famoso « postulato di Kock » è stato soddisfatto, e il tumore si può decisamente collocare tra quelli trasmessi da un agente infettivo.

I primi esperimenti sui tumori della rana sono serviti come importanti precursori del lavoro attuale sulle cause del cancro nell'uomo. I risultati ottenuti hanno sollevato l'interesse di altri ricercatori circa le prospettive di scoprire tumori umani che abbiano origine da virus erpetici.

Una formazione tumorale maligna del sistema linfatico dei polli, nota come malattia di Marek, aveva portato alla prima inequivocabile prova che un virus erpetico è la causa di un tipo di cancro. Quasi sconosciuta qualche anno fa, eccetto che tra un piccolo gruppo di patologi veterinari accademici, la malattia di Marek divenne, verso la fine degli anni 60, il flagello del commercio del pollame negli Stati Uniti. In un primo tempo, la malattia che colpiva i polli veniva chiamata polineurite o paralisi del pollo. In seguito fu diagnosticata come la malattia che era stata descritta nel 1907 dall'ungherese Joseph Marek. Marek non aveva tuttavia compreso che la malattia era di tipo maligno. Le cellule del sistema linfatico diventano cancerose e invadono le cellule nervose, con il risultato della paralisi del pollo. Man mano che la malattia avanza vengono colpiti anche altri organi, e alla fine il pollo muore.

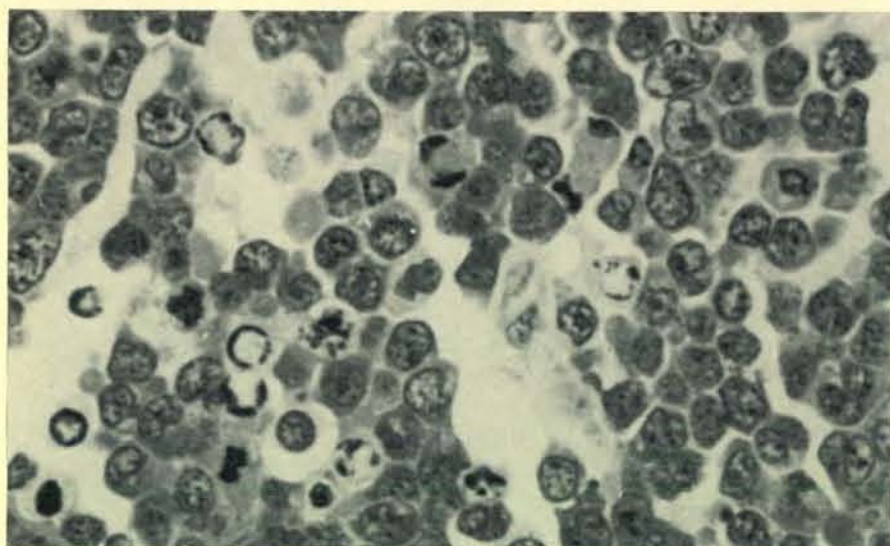
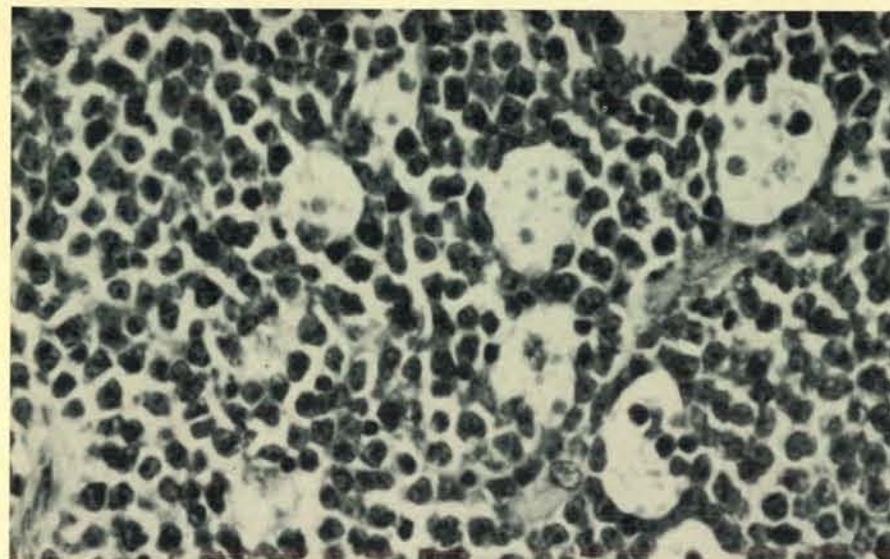
Nel 1967, numerosi ricercatori dimostrarono contemporaneamente che quando un omogenato filtrato proveniente da un pollo infetto, veniva iniettato in un pollo sano, si sviluppava un tumore linfoide. Sebbene in alcuni tumori del tessuto sia stata osservata la presenza di virus, tali osservazioni erano rare. Questo aprí il problema su come il virus fosse trasmesso da un pollo all'altro.

Keyvan Nazerian del Laboratorio regionale di ricerca sul pollame del Dipartimento per l'agricoltura di East Lansing nel Michigan, usando un test antigene-anticorpo, esaminò le cellule da diversi tessuti di polli con la malattia di Marek nel tentativo di determinare come e dove il virus venisse liberato nell'ambiente. Le cellule tumorali contenevano raramente un virus, ma quando Nazerian esaminò delle sezioni di pelle, trovò che le cellule dei follicoli delle penne davano risultato positivo per l'esame della presenza del virus. Fu presto dimostrato che le penne del pollo e le desquamazioni della pelle sono molto infette e sembra chiarito il modo di trasmissione del virus nei gruppi di polli commerciali. Il virus fu identificato come un virus erpetico. Un interessante risultato di questo lavoro è la scoperta che i virus erpetici non sono necessariamente delicati come invece la maggior parte dei virus. Il virus che provoca la malattia di Marek è oltremodo resistente all'essiccamento e può rimanere infettivo per un lungo periodo dopo che è stato liberato nelle desquamazioni del pollo.

Sono stati preparati dei vaccini con virus vivi, impiegando un ceppo attenuato del virus della malattia di Marek e un virus simile preso dai tacchini, con il risultato che la malattia di Marek è stata efficacemente sradicata nei gruppi di polli trattati. Questo suggerisce che per l'uomo sia possibile preparare dei vaccini contro il cancro. Un punto utile da osservare è che nella malattia di Marek il virus tumorale non viene prodotto nelle cellule tumorali. Il virus trasforma cellule normali in cellule tumorali senza necessariamente produrre una progenie virale. In effetti, i tumori che producono nuovi virus erpetici sarebbero probabilmente autolimitanti, poiché si pensa che la produzione del virus provochi la morte della cellula implicata. È un fatto interessante che la replicazione del virus erpetico della malattia di Marek avviene nelle cellule cutanee non maligne che di solito muoiono in ogni caso.

Si è visto che altri tumori degli animali sono causati da virus erpetici o sono con essi strettamente collegati. Uno di questi è la adenomatosi polmonare della pecora, un tumore del polmone che è responsabile di notevoli perdite commerciali tra gli allevatori di pecore. Harry C. Hinze della Scuola di medicina dell'Università del Wisconsin ha trovato che il virus erpetico del coniglio era in grado di provocare dei tumori linfatici maligni quando veniva iniettato in conigli giovani. Entrambe queste associazioni virus-tumore sono ancora allo studio.

I virus erpetici che provocano il tumore di Lucké e il tumore della malattia di Marek sono entrambi connessi, da un punto di vista immunologico, con un virus erpetico che è strettamente associato con un cancro umano: il linfoma di Burkitt. Nel 1959 Denis Burkitt, un chirurgo missionario in Uganda, osservò che un gran numero di bambini africani, in età dai quattro ai 16 anni, era affetto da un tumore nel tessuto connettivo della mascella. Si pensava che questi sarcomi della mascella fossero poco frequenti in Europa e negli Stati Uniti, ma i figli di missionari bianchi che risiedevano in Uganda venivano qualche volta colpiti dalla malattia. Burkitt osservò anche che tutti i casi si verificavano in zone dove la malaria era endemica, o in zone dove le condizioni atmosferiche erano tali da far pensare

Il linfoma di Burkitt in Africa è frequente in quelle zone (*punti colorati*) dove i fattori climatici rendono endemica la malaria. In tali zone (*colore chiaro*) l'altitudine è inferiore ai 1500 metri, la temperatura media stagionale è di 15,6 gradi °C o anche piú e la pioggia raggiunge o supera i 500 millimetri all'anno. Il linfoma di Burkitt è un sarcoma umano della mascella strettamente connesso con un virus erpetico.

Il tumore del linfoma di Burkitt in sezione microscopica (*in alto*) mostra il caratteristico aspetto a «cielo stellato» di tali tumori. Le «stelle» bianche sono grandi istiociti, cellule che fagocitano residui di cellule morte. Il materiale incorporato è visibile all'interno degli istiociti. Un ingrandimento maggiore delle cellule tumorali (*in basso*) rivela le strutture circolari (*inclusioni intranucleari di tipo A*) che sembrano vuote e che sono caratteristiche delle infezioni da virus erpetico. Le microfotografie sono state fornite da D.H. Wright dell'Università di Birmingham.

che fossero presenti in gran numero le zanzare che trasmettono la malaria (*si veda la figura a pagina 23*).

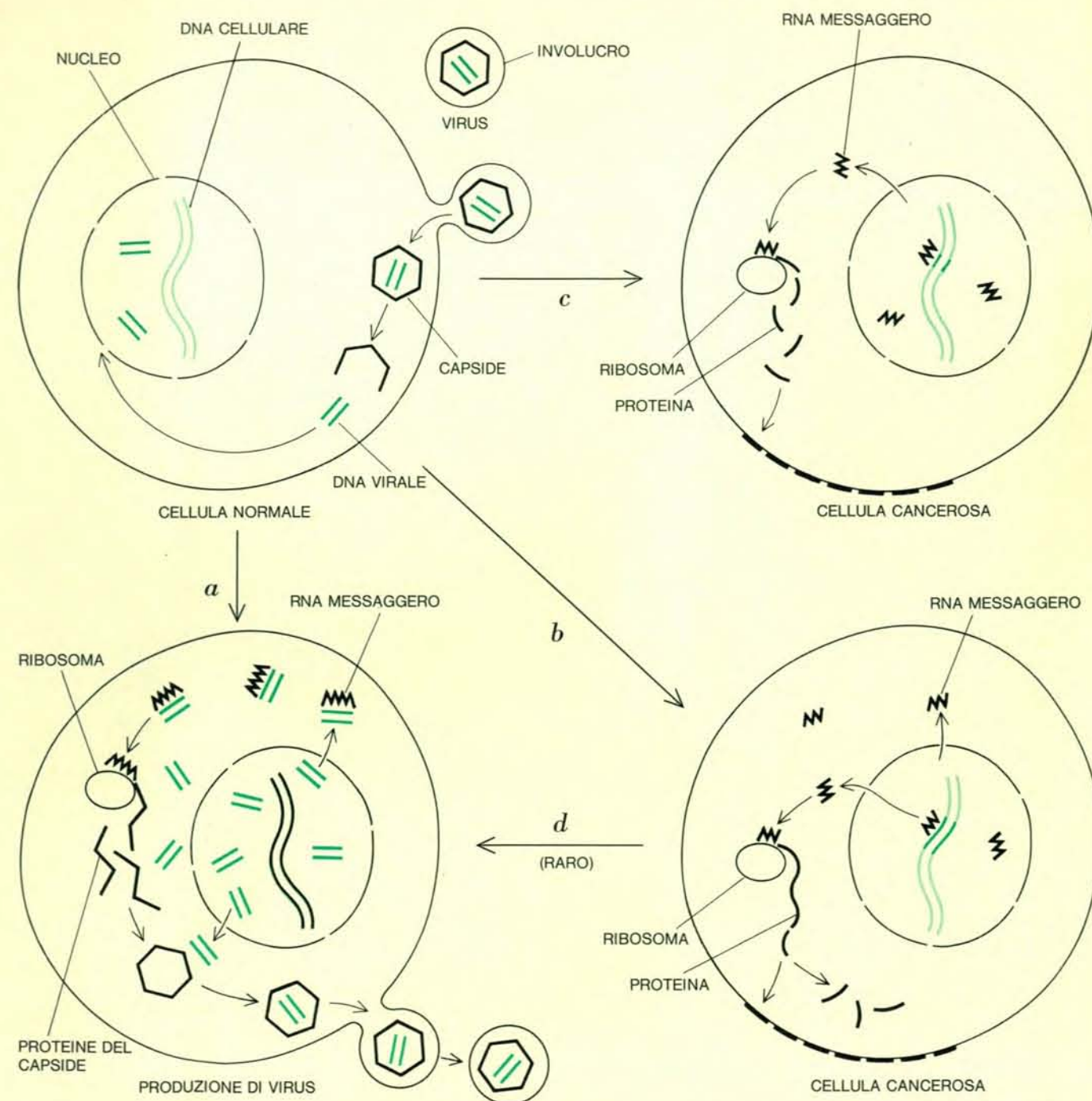
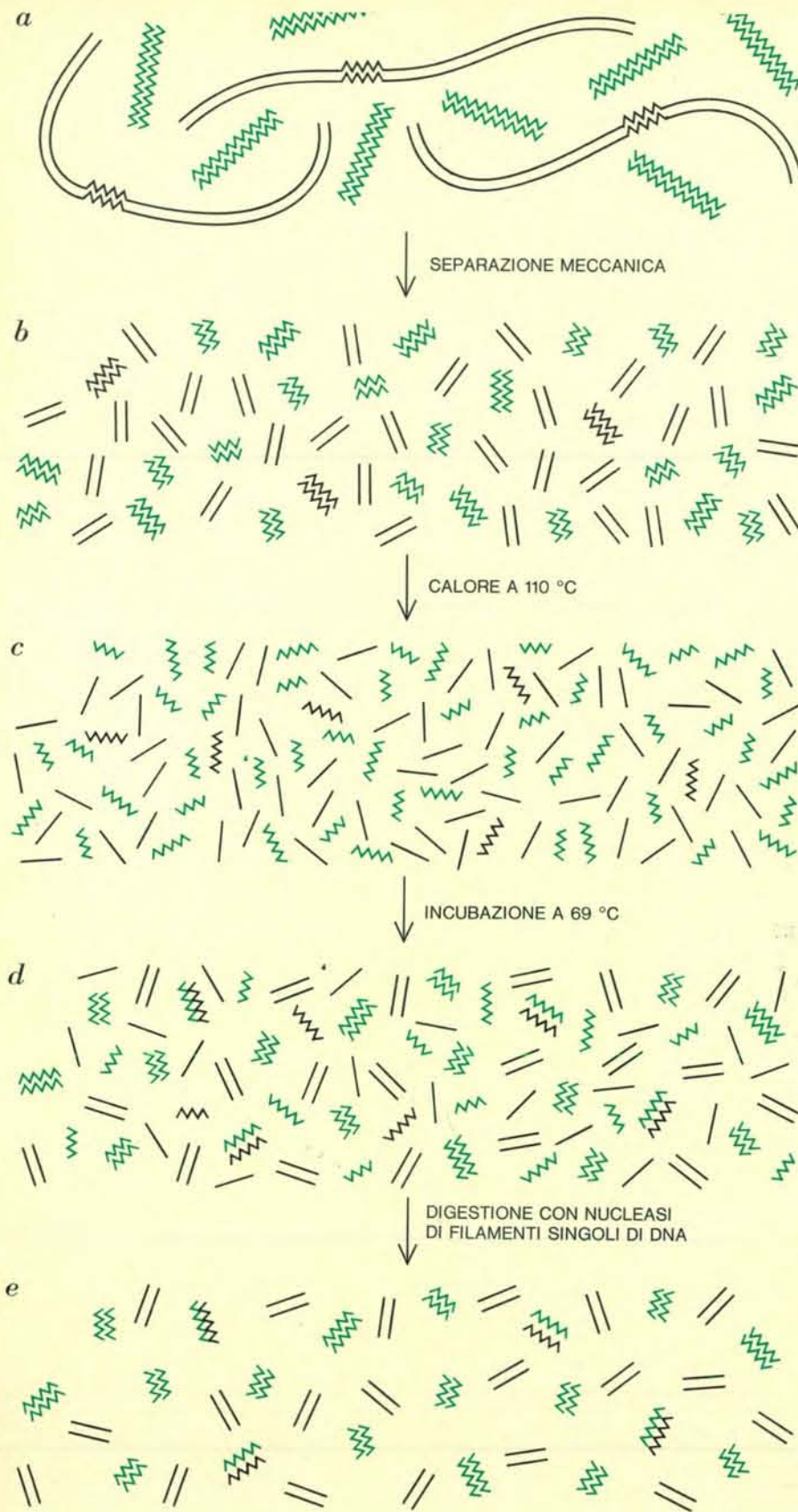
Intensi studi sul linfoma di Burkitt cominciarono immediatamente dopo la pubblicazione del suo rapporto su questo tipo di cancro che sembrava venisse trasmesso come una malattia infettiva. Seguirono presto numerose scoperte. Si trovò che il tumore è raro nei bambini africani di età superiore ai 16 anni. Una possibile spiegazione è che il contatto con qualche agente infettivo sia frequente in Uganda, cosicché tutti i nati in una certa zona vi siano esposti nei primi anni di vita. Inoltre il fenomeno di concentrazione temporale della malattia, sospettato sulla base dei ricoveri in ospedale, è stato confermato da M. Pike dell'Università di Oxford e R.H. Morrow jr. Per esempio, nella Contea del Bwamba nell'Uganda, non si sono verificati casi di linfoma di Burkitt prima del 1966; in seguito furono riportati 7 casi in 27 mesi e da allora non ne è stato piú riportato alcuno. Scoperte di questo tipo indicano chiaramente che la malattia viene trasmessa da un agente infettivo. In seguito K. Booth, Burkitt e i loro colleghi trovarono che il linfoma si ha anche tra la gente della Papuasia e della Nuova Guinea. In entrambi i luoghi le condizioni climatiche sono tali per cui la malaria è endemica.

Sebbene la mascella inferiore sia il luogo piú frequente dove si sviluppa il linfoma di Burkitt, i tumori insorgono anche in altre zone: la mascella superiore, la tiroide, le ovaie, il fegato e i reni.

D.H. Wright, che lavorava allora all'Università Makerere in Uganda, fece lunghi studi al microscopio sul tumore. Egli trovò che si origina in cellule dei linfonodi. Sebbene il tumore sia maligno, in quanto invade e distrugge i tessuti vicini (e infine causa la morte del paziente se non viene curato), le cellule del sangue circolante sembrano rimanere normali, almeno fino a quando la malattia non raggiunge uno stadio avanzato. Questo fatto costituisce la differenza tra i linfomi e la leucemia, dove le cellule del sangue diventano immediatamente anormali.

Un'altra scoperta fatta da Wright era che le sezioni microscopiche di tumori prelevate dal linfoma di Burkitt, contengono istiociti, grandi cellule chiare che incorporano frammenti provenienti da cellule morte della massa tumorale. È questa scoperta che ha aiutato altri ricercatori a identificare casi di linfoma di Burkitt in parti del mondo dove esso è piú raro e dove con meno probabilità si sviluppa nella parte piú caratteristica: la mascella inferiore. Ci furono dei risultati inattesi con i nuovi metodi diagnostici, e si cominciarono ad avere rapporti secondo i quali il linfoma di Burkitt insorge regolarmente nel Nord e Sud America, e in tutta l'Europa, compresi i paesi piú freddi come la Scandinavia, dove la malaria trasmessa dalla zanzara è praticamente sconosciuta. Quindi sembra che il linfoma di Burkitt venga in effetti trasmesso come una malattia infettiva, ma non necessariamente dalle zanzare. Per mantenere le cose nella giusta prospettiva si deve sottolineare che anche in Uganda, dove forse la malattia è piú diffusa, la probabilità che essa si sviluppi è di un caso su 2000. Nelle zone temperate la probabilità è molto piú bassa.

Una volta che si ebbe il sospetto che la causa del linfoma di Burkitt fosse un virus, furono fatti numerosi sforzi per dimostrarne la presenza nel tumore. La ricerca, tuttavia, si rivelò difficile. L'esame al microscopio elettronico di sezioni di tumore per rilevare le particelle del virus fu inutile, lo stesso dicasi per i tentativi di infettare animali e colture di tessuti con estratti del tumore. All'inizio, in realtà, anche i tentativi di coltivare le stesse cellule tumorali in colture di tessuto ebbero scarso successo. In seguito, M.A. Epstein e Y.M. Barr della Scuola di medicina dell'Università di Bristol cominciarono a lavorare con tessuto tumorale fresco che arrivava in volo dall'Uganda e R.J.N. Pulvertaft, che lavorava allora all'Università di Ibadan in Nigeria cominciò esperimenti simili. Epstein e Barr trovarono che le cellule crescevano in due delle loro colture, ma che ciò avveniva solamente in sospensione, e non aderendo alla superficie del recipiente di coltura, dove normalmente crescono le cellule in colture di tessuto. Da un punto di vista retrospettivo, poiché le cellule del linfoma e altre cellule dei linfonodi circolano periodicamente nel sangue, ci si sarebbe dovuti aspettare che le cellule crescessero in sospensione. Quando le cellule della coltura di Epstein-Barr furono esaminate al microscopio elettronico, si poterono vedere chiaramente le particelle del virus erpetico. Anche Pulvertaft riuscí a dimostrare la presenza di un virus erpetico nelle sue colture di tessuto.

Epstein e Barr dimostrarono che le cellule del linfoma potevano essere

L'invasione di una cellula da parte di un virus erpetico porta normalmente alla produzione di nuove particelle virali. Il DNA virale si replica all'interno del nucleo della cellula. Poi, una parte si sposta verso il citoplasma dove trascrive l'RNA messaggero che si attacca ai ribosomi (*a*). Le proteine dell'involucro virale sono riunite in capsidi, che si combinano con il DNA per produrre una particella virale completa. La cellula muore quando le particelle virali vengono liberate. In alcuni casi (*b*) il DNA virale completo si combina con il DNA cellulare. L'RNA messaggero viene trascritto e quando questo si attacca ai ribosomi, produce delle proteine che si pensa migrino verso la superficie della cellula. Questi aggregati proteici sulla superficie della cellula potrebbero essere implicati nella sua trasformazione in una cellula cancerosa. In rari casi, il DNA virale in una cellula cancerosa può tornare a produrre virus (*d*). Un'altra via possibile per indurre il tumore (*c*) si può avere quando soltanto parte del DNA virale si combina con il DNA della cellula. In questo caso non si possono formare nuovi virus completi. L'RNA messaggero, tuttavia può essere trascritto, permettendo la sintesi di proteine nelle cellule tumorali.

La tecnica di ibridazione permette ai ricercatori di misurare la quantità di DNA virale (o RNA) presente nel DNA cellulare. Il DNA virale radioattivo (*segmenti colorati*) è mescolato con il DNA cellulare (*a*). Le molecole vengono spezzate facendo passare la miscela attraverso una stretta apertura. I segmenti corti di DNA che ne risultano (*b*) vengono riscaldati per trasformare i doppi filamenti in filamenti singoli (*c*). Quando la miscela si raffredda, si riformano alcuni doppi filamenti permettendo al DNA virale radioattivo di formare ibridi con qualsiasi segmento virale nei segmenti di DNA cellulare (*d*). La miscela è trattata con un enzima che digerisce i filamenti singoli di DNA (*e*); misurandone quindi la radioattività si ottiene la quantità di DNA ibrido.

propagate indefinitamente nelle colture di tessuto e che una delle colture (chiamata EB-3) continua a produrre un virus infettivo. Soltanto circa il 2 per cento delle cellule muore ogni tanto e libera il virus. Altri ricercatori hanno dimostrato che il virus, ora chiamato il virus di Epstein-Barr, o EBV, può infettare cellule del sangue normali in colture e indurle alla produzione del virus.

L'identificazione di un virus erpetico nel linfoma di Burkitt fu significativa perché è noto che alcuni virus erpetici sono molto diffusi tra la popolazione umana. Nel 1968 Werner e G. Henle e V. Diehl della Scuola di medicina dell'Università di Pennsylvania scoprirono che la mononucleosi infettiva, una malattia a decorso benigno nota come la « malattia da bacio » tanto comune tra gli studenti delle Università americane ed europee, è causata da un virus erpetico che non può essere distinto dal virus di Epstein-Barr, neppure usando i più sofisticati test di ibridizzazione dell'acido nucleico o antigene-anticorpo. Quindi, i due virus sono presumibilmente identici. Questa interessante scoperta fu fatta per caso: si trovò che una tecnica che lavorava nel laboratorio degli Henle aveva sviluppato anticorpi contro il virus di Epstein-Barr dopo aver contratto la mononucleosi infettiva. Gli Henle seguirono questa traccia e dimostrarono che quando delle cellule del sangue di una persona con mononucleosi infettiva sono coltivate in una coltura in sospensione, esse producono un virus che non è distinguibile dal virus di Epstein-Barr.

Il sangue di pazienti con mononucleosi infettiva contiene cellule che si dividono e che non si riscontrano mai nel sangue di persone non infettate. La condizione è stata, infatti, descritta come una malattia autolimitante. L'esame al microscopio delle cellule in divisione rivela che esse sono molto simili alle cellule bianche in proliferazione di casi di leucemia linfoblastica acuta. Inoltre, le sezioni microscopiche di linfonodi di pazienti con mononucleosi infettiva sono notevolmente simili, per certi aspetti, alle sezioni di tumori del linfoma di Burkitt.

Le persone che hanno avuto la mononucleosi infettiva hanno una maggiore probabilità di contrarre il cancro più avanti negli anni? Studi epidemiologici intrapresi da Robert W. Miller dell'Istituto nazionale del cancro e da Gilbert W. Beebe della Accademia nazionale delle scienze (Consiglio nazionale delle ricerche) indicano che precedenti di mononucleosi infettiva non sembrano essere associati con l'incidenza del cancro. Questo dato è in accordo con il punto di vista di molti, per cui praticamente tutti sono infettati dal virus di Epstein-Barr, compresi quelli che non hanno mai contratto la mononucleosi.

Ci sono molte malattie del sangue che sembrano essere tra le principali candidate per quanto riguarda la connessione con i virus erpetici, o per le quali essi possono essere perfino la causa. Si è trovato che i pazienti affetti dalla malattia di Hodkin, che colpisce il sistema linfatico, presentano quasi sempre alti livelli di anticorpi per il virus di Epstein-Barr. Anche i pazienti affetti da leucemia linfocitica acuta tendono a mostrare elevati livelli di anticorpi verso il virus di Epstein-Barr. Nella malattia di Hodgkin c'è una proliferazione maligna delle cellule del reticolo, le quali, viste al microscopio, sono molto simili alle cellule istiocitarie della mononucleosi infettiva e alle caratteristiche cellule « a cielo stellato » del linfoma di Burkitt. Un altro cancro associato con il virus di Epstein-Barr è il carcinoma nasofaringeo. Lloyd V. Old e i suoi colleghi dell'Istituto Sloan-Kettering fecero questa scoperta quando usavano sieri di pazienti con carcinoma nasofaringeo come controlli in studi antigene-anticorpo di sieri di pazienti con il linfoma di Burkitt.

Come indica il termine, il carcinoma nasofaringeo ha origine nella cavità nasale, la faringe e la parte posteriore della cavità orale. È un vero carcinoma in quanto il tumore consiste di cellule epiteliali proliferanti in superficie, piuttosto che di cellule linfoidi di tessuti profondi. Fin quando venne fatta questa scoperta, al virus di Epstein-Barr erano stati associati soltanto tumori di tessuti profondi. È particolarmente interessante notare che una struttura prominente della faringe superiore sono le tonsille, che sono composte di cellule linfoidi ricoperte da una superficie epiteliale. Questo epitelio è perforato, e il materiale nella bocca viene a contatto con il tessuto tonsillare. È chiaro che le cellule linfoidi si possono infettare e trasmettere l'infezione alle cellule epiteliali. Inoltre, i prelievi provenienti dalla gola di pazienti affetti da mononucleosi infettiva sono un'importante fonte del virus di Epstein-Barr. L'incidenza del carcinoma nasofaringeo è straordinariamente alta in certe popolazioni. Praticamente tutti i casi noti sono stati osservati in popolazioni cinesi della Cina meridionale, della Malesia e di Singapore. H.C. Ho del Queen Elizabeth Hospital di Hong Kong ha riunito dei dati che suggeriscono che l'incidenza media a Hong Kong può arrivare fino ad 1 caso su 200. Tra « gli abitanti delle barche » di Hong Kong l'incidenza è anche superiore. In persone di discendenza cinese nate negli Stati Uniti, l'incidenza della malattia è minore, ma è ancora alta rispetto all'incidenza nel resto della popolazione. Come il linfoma di Burkitt, anche il carcinoma nasofaringeo è diffuso in tutto il mondo occidentale, ma con un'incidenza bassa. Ciononostante i dati sono ancora inconcludenti; essi fanno ancora pensare che accanto a un possibile fattore virale possa essere implicato un fattore legato a determinate abitudini.

Guy de-Thé del gruppo di ricerca sul cancro della Organizzazione mondiale della sanità a Lione ha coltivato in sospensione cellule provenienti da tessuto tumorale di carcinoma nasofaringeo. Questa coltura produce particelle di virus che sembrano essere identiche a quelle del linfoma di Burkitt, ma ulteriori studi sono necessari in proposito.

La forma genitale dell'*Herpes simplex* (Tipo 2) si presenta con ulcere o macchie arrossate sulle pareti della cervice e della vagina di donne infette e sul pene e nel tratto genitale degli uomini. Isolina M. Centifanto e i suoi colleghi del Collegio di medicina dell'Università della Florida hanno trovato che il 15 per cento degli uomini di tutte le età, dai 15 anni in sú, ospitano l'*Herpes simplex* di Tipo 2 nel tratto genitale, per lo piú nella prostata e nei vasi deferenti. Il virus viene raramente isolato dal tratto genitale femminile; le donne tendono ad avere infezioni genitali da virus erpetico acute, di breve durata e ricorrenti. È possibile che molti uomini ospitino il virus senza sviluppare i sintomi dell'infezione, ma possano trasmetterla.

André J. Nahimias, della Scuola di medicina della Emory University ha potuto stabilire una valida associazione tra il virus *Herpes simplex* di Tipo 2 e il cancro della cervice, e la Centifanto ha descritto un caso in cui il virus erpetico poteva essere messo in evidenza nelle cellule di tessuto proveniente da un cancro della prostata. Prove di un certo interesse vengono da uno studio intrapreso da Niza Frenkel insieme con i suoi collaboratori dell'Università di Chicago, i quali lavoravano con del tessuto tumorale fornito da Nahimias. Lo studio riguardava esperimenti di ibridazione con DNA proveniente da cellule di carcinoma del collo uterino. Trovarono che in queste cellule è effettivamente presente DNA proveniente da virus erpetico del Tipo 2, ma che il 60 per cento della molecola del DNA virale manca. La parte mancante può spiegare per quale motivo non sia possibile isolare dal tessuto tumorale né il virus completo, né gli antigeni virali. La porzione del DNA virale che è presente potrebbe attaccarsi al DNA della cellula normale e potrebbe essere responsabile del mantenimento della cellula in una condizione tumorale. Inoltre la parte mancante sarebbe necessaria per la produzione del virus infettante completo.

Di grande importanza è la recente scoperta fatta da Luis V. Melendez del Centro regionale di ricerca sui primati del New England secondo cui i saimiri scoiattolo americani ospitano un virus erpetico che è in grado di indurre tumori nella mirichina o aoto e nell'uistiti. È stato possibile, per la prima volta, fare esperimenti di trasmissione del cancro su animali molto piú vicini all'uomo rispetto alle rane e ai polli. Ancora piú promettente è il fatto che i tumori che si originano nelle scimmie sono simili ai tumori associati all'*Herpes* dell'uomo. A differenza delle cellule di coltura infettate col virus di Epstein-Barr, le cellule che ospitano il virus del saimiri scoiattolo si sviluppano sulla superficie del recipiente di coltura. Questo rende piú semplice la produzione di grandi quantità di virus per la ricerca o per la preparazione di vaccini.

Abbastanza recentemente, due gruppi di ricercatori hanno dimostrato che il virus di Epstein-Barr può indurre tumori linfoidi nelle scimmie. Uno di questi ricercatori è stato lo stesso Epstein che utilizzava il virus per indurre il tumore in una mirichina. Thomas C. Shope, D.C. De Cahiro e I. George Miller jr. della Scuola di medicina dell'Università di Yale, hanno provocato dei linfomi maligni in 4 uistiti. Poiché tali esperimenti non possono essere fatti sull'uomo, questo è il massimo che possiamo fare per dimostrare direttamente che un virus proveniente da un cancro umano è effettivamente un virus cancerogeno, cosa che però è ancora ben diversa dal trovare un virus in un tumore.

Ci sono quindi buone speranze di poter dimostrare infine, al di là di ogni dubbio, che i virus erpetici possono causare alcune forme di cancro nell'uomo. Tuttavia non si può escludere la spiegazione alternativa che i virus erpetici possano trovarsi in alcuni tumori umani solo come ospiti e non essere implicati nello sviluppo del tumore stesso.

Conformazione delle proteine e controllo biologico

I processi vitali sono messi in moto e interrotti da un meccanismo di controllo universale che dipende dalla capacità delle molecole proteiche di assumere conformazioni diverse a seconda delle influenze esterne

di Daniel E. Koshland jr.

Considerando il gran numero di processi che devono essere regolati da un organismo vivente e la varietà delle condizioni ambientali alle quali esso deve reagire, verrebbe fatto di pensare che i meccanismi di controllo, nei sistemi biologici, siano straordinariamente complessi. Questo è vero; tuttavia, se si esaminano i processi piú da vicino, ci si accorge che gli elementi fondamentali dei meccanismi di controllo sono in realtà molto semplici e universali.

L'elemento fondamentale del meccanismo di controllo in tutti i sistemi viventi – dal piú piccolo batterio all'uomo – è la molecola proteica. Gli enzimi, i catalizzatori biologici che controllano tutti i processi chimici dei sistemi viventi, sono proteine. I recettori sensoriali, che ci pongono in grado di vedere, di udire, di gustare, di odorare, sono proteine. Gli anticorpi che ci immunizzano contro le infezioni, sono proteine. Recenti esperimenti hanno dimostrato che è la capacità di queste proteine di cambiare forma per effetto delle influenze esterne a fornire il meccanismo di controllo cosí essenziale per il sistema vivente.

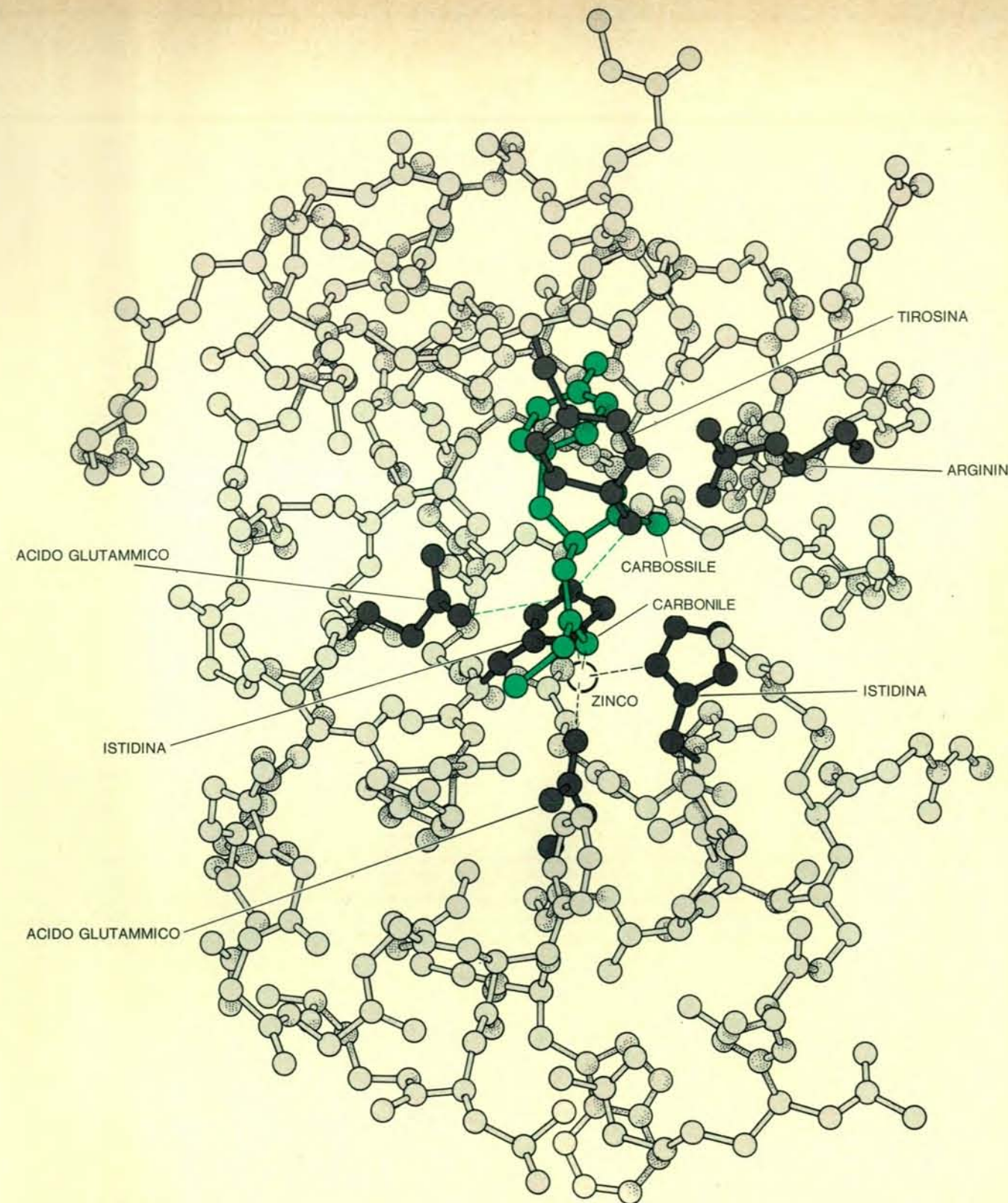
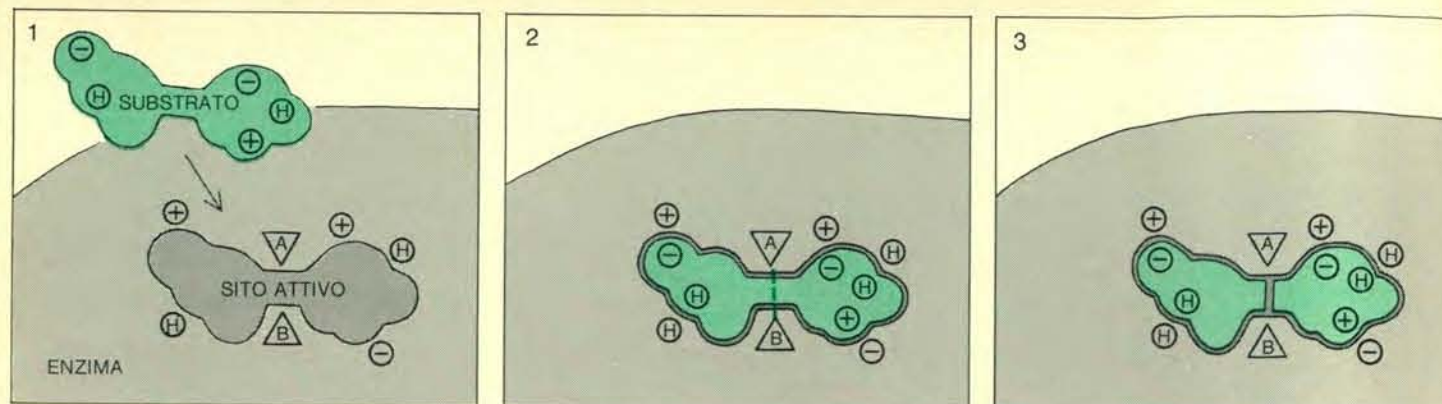
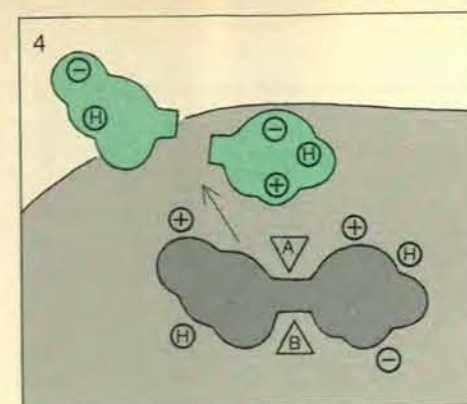
Il punto di partenza per l'elaborazione di questo concetto è stato lo studio degli enzimi, il che non sorprende se si considera che gli enzimi sono le molecole piú facili da studiare. È stato calcolato che una cellula vivente di media grandezza contiene qualcosa come tremila differenti enzimi. Ciascuno di essi catalizza una specifica reazione chimica nella quale dei composti, chiamati substrati, vengono trasformati in altri composti chiamati prodotti. Fortunatamente per la nostra comprensione dei sistemi biologici, molti enzimi sono molecole piuttosto resistenti che è possibile estrarre da un sistema fisiologico senza che le loro proprietà biologiche vadano distrutte. Possono quindi essere studiati in provetta ed essere posti in condizione di svolgere la medesima funzione catalizzatrice che hanno nell'organismo vivente.

Inoltre è possibile riprodurre nella provetta le stesse condizioni ambientali con cui vengono a contatto nella cellula vivente, il che permette di dedurre il ruolo da essi svolto nella regolazione biologica.

Modificazioni di forma e catalisi enzimatica

Era noto da tempo che il meccanismo fondamentale per mezzo del quale gli enzimi catalizzano le reazioni chimiche comincia con la formazione di un legame tra il substrato (o i substrati) e la superficie dell'enzima; l'enzima, a questo punto, polarizza i legami chimici nel substrato causando una reazione che porta alla formazione di prodotti sulla superficie. Il distaccarsi di questi prodotti dalla superficie rigenera l'enzima libero e permette un nuovo ripetersi del ciclo.

Diversamente dai catalizzatori preparati in laboratorio, gli enzimi hanno un'importante caratteristica che va sotto il nome di specificità; ciò vuol dire che uno solo, o tutt'al piú pochissimi composti chimici possono reagire con un particolare enzima. Questa proprietà può essere spiegata dall'ipotesi della « serratura », formulata nel 1894 da Emil Fisher, secondo la quale l'enzima avrebbe una forma tale che solo alcuni composti possono aderire alla sua superficie, cosí come si adattano tra loro una chiave e una serratura o due elementi di un puzzle. La complementarità delle forme permette che un composto si attacchi all'enzima escludendo altri composti che non hanno la giusta dimensione, forma o distribuzione delle cariche (*si veda la figura alle pagine 30 e 31*). La moderna cristallografia a raggi X ha rivelato fin nei dettagli un'aderenza proprio di questo tipo fra enzima e substrato (*si veda l'illustrazione nella pagina a fronte*).

Sebbene questa ipotesi potesse spiegare molti dati inerenti alla specificità, tuttavia sono emerse alcune vistose discrepanze. Per esempio, si è visto che alcune sostanze, di dimensioni o troppo grandi o troppo piccole, si legavano con la superficie dell'enzima, pur senza dar luogo a composti. Inoltre era difficile, sulla base della teoria della sagoma rigida, spiegare come gli zuccheri potessero competere con l'acqua nelle reazioni enzimatiche e perché il legame dei substrati avvenisse solo in un determinato ordine di successione. Questi fatti, e altri analoghi, hanno portato a formulare l'ipotesi che l'enzima non possieda sin dall'inizio una forma complementare a quella del substrato, ma che esso venga piuttosto indotto ad assumere la forma complementare cosí come una mano produce un cambiamento nella forma di un guanto (*si veda l'illustrazione alle pagine 34 e 35*). Questa teoria della « complementarità indotta » presume che il substrato giochi una sua parte nel determinare la forma finale dell'enzima e che l'enzima stesso sia quindi flessibile. La dimostrazione dell'effettivo cambiamento di forma delle proteine per influenza di piccole molecole fu ottenuta dapprima mediante studi biochimici che dimostrarono come le catene laterali di amminoacidi allineate lungo l'asse centrale della molecola proteica reagiscano in modo diverso a certe sostanze proteiche. L'ipotesi è stata in seguito confermata con l'ausilio delle tecniche fisiche piú avanzate, in particolar modo della cristallografia a raggi X. Una volta accettato il concetto di enzima

La precisa aderenza tra il sito attivo sulla superficie di una grossa molecola proteica e la molecola del substrato specifico con cui la proteina reagisce è evidente in questo disegno semplificato a tre dimensioni, basato su dati ottenuti mediante cristallografia a raggi X da William N. Lipscomb jr. e i suoi colleghi dell'Università di Harvard. La proteina, in varie gradazioni di grigio, è la carbossipeptidasi A, un enzima digestivo la cui funzione, come dice il nome, consiste nel rompere la catena polipeptidica del substrato vicino al suo carbossile terminale. Il substrato, in colore, è carbobenzossilalanilalanil-tirosina. In questa figura è indicato circa un quarto del numero totale di atomi delle catene polipeptidiche contenute nelle due molecole. Gli atomi illustrati sono per la maggior parte atomi di carbonio, con una piccola aggiunta di atomi di azoto e di ossigeno; tutti gli atomi di idrogeno sono stati omessi. Le sei catene laterali del sito attivo che interagiscono in modo specifico con il substrato sono in grigio scuro. Per esempio, si vede che una catena laterale di arginina a carica positiva attrae il gruppo carbossilico a carica negativa del substrato. Inoltre certe regioni idrofobe, od oleose, del substrato sono attratte da regioni simili sull'enzima, sicché l'attrazione fra le due molecole viene a essere rafforzata. Un atomo di zinco (*in bianco*) costituisce un ulteriore « legame di coordinazione » che coinvolge un gruppo carbonile del substrato e tre altre catene laterali di amminoacidi (acido glutammico e due istidine) che si espandono dall'interno della concavità del sito attivo dell'enzima. La catena laterale di tirosina e la seconda catena laterale di acido glutammico del sito attivo sono gruppi catalitici che polarizzano gli elettroni in uno dei legami chimici del substrato, rompendo questo legame (*linea spezzata in colore*) e cosí dividono il substrato in due parti che costituiscono i prodotti della reazione di questa particolare combinazione enzima-substrato.

Modello «a serratura» del meccanismo con cui gli enzimi catalizzano le reazioni chimiche, ideato da Emil Fisher nel 1894. Il sito attivo dell'enzima è una struttura sagomata rigida che consente di aderire solo a speciali composti, cosí come una chiave si adatta a una serratura. Un composto di questo tipo, chiamato substrato, viene attratto verso il sito attivo dell'enzima da gruppi forniti di attrazione reciproca, come le cariche elettrostatiche su certe catene laterali di amminoacidi (+, —), e dalla tendenza alla coalescenza di catene laterali idrofobe adiacenti (*H*). Nel modello a serratura i gruppi catalitici (*A*, *B*) vengono posti in equilibrio anticipatamente (*1*) perché avvenga la reazione che rompe il legame chimico nella strozzatura

del substrato (*2*), trasformando il substrato nei composti chiamati prodotti (*3*). Il distacco di questi composti dalla superficie (*4*) rigenera l'enzima libero e consente la ripetizione del ciclo. Secondo la teoria qui esposta la precisa complementarità delle forme del substrato e dell'enzima che su questo agisce impedisce la reazione con altri composti che abbiano dimensioni o troppo grandi per poter entrare in contatto con il sito attivo (*5a*), o dimensioni troppo piccole per venire da questo attratti (*5b*).

flessibile, divenne possibile spiegare lo strano comportamento di alcune classi particolari di composti che si legavano con l'enzima senza dar luogo a prodotti. Un tipo di molecola troppo grande, per esempio, aderisce alla superficie dell'enzima, ma nel far ciò induce nella proteina una distorsione che non consente ai gruppi catalitici di allinearsi correttamente. Anche altri composti possono legarsi con l'enzima, ma ad essi fanno difetto o la giusta dimensione oppure le caratteristiche chimiche necessarie per indurre il corretto allineamento. Pertanto, anche se entrambi i tipi di composti si legano con l'enzima, nessuno reagisce con esso.

Modificazioni di forma e regolazione

La scoperta della flessibilità non comporta che tutte le proteine debbano essere flessibili; alcune, al contrario, possono essere assolutamente rigide, e per queste ultime la teoria della serratura offre una spiegazione validissima. Non significa neppure che tutti gli enzimi flessibili debbano comportarsi come tali con tutti i composti chimici. Tale scoperta significa comunque che la flessibilità della proteina è una fondamentale caratteristica dell'azione dell'enzima. È stato comunque dimostrato che la capacità di produrre una modificazione di forma costituisce una caratteristica vitale e diffusa della maggior parte, se non di tutti gli enzimi.

Il concetto di flessibilità delle proteine indusse a ritenere che piccole molecole, di per sé non coinvolte nella reazione chimica, potessero aiutare a far agire da substrato una molecola non adatta alterando la forma dell'enzima (*si veda l'illustrazione a pagina 32*). Nel caso di una molecola troppo piccola per portare al corretto allineamento dei gruppi catalitici, per esempio, può essere utile l'aggiunta di alcune molecole che, pur non entrando nella reazione, permettono di ottenere una forma stabile che presenta il corretto allineamento dei gruppi catalitici. Un modo per ottenere questo nuovo allineamento è che una seconda molecola si leghi in posizione immediatamente adiacente alla molecola insufficiente, dando cosí al sito attivo la forma appropriata.

L'ipotesi della flessibilità dell'enzima è stata in seguito confermata in molti casi (per esempio per l'enzima esochinasi, che scinde i carboidrati, da Alberto Sols e i suoi collaboratori in Spagna e per l'enzima tripsina da T. Inagami e T. Murachi in Giappone). Anche molecole che si legano in posizione molto lontana dal sito catalitico possono indurre una forma appropriata. In questo caso la complementarità indotta si trasmette attraverso la proteina – come un urto impresso al primo pezzo di una fila di domino li fa cadere tutti –, fino a quando il sito attivo non è stato adeguatamente alterato.

Può anche darsi il caso inverso; gli enzimi flessibili possono essere indotti ad assumere una forma non attiva da molecole dette inibitori. Queste molecole possono alterare nell'enzima o la funzione catalitica o la capacità di legarsi e in entrambi i casi si ha come risultato una forma inattiva (*si veda l'illustrazione a pagina 33*).

In breve, una molecola regolatrice che non è di per sé coinvolta nella reazione chimica, può controllare l'attività di un enzima modificandone la forma. Inducendo la forma appropriata essa dà il via all'attività enzimatica, facendo il contrario la interrompe. Nei sistemi biologici uno dei gruppi piú importanti di molecole di questo tipo è costituito dagli ormoni. Sebbene gli ormoni vengano prodotti in piccola quantità, troppo piccola per conferire loro un'importanza diretta come riserve alimentari o fonti di energia, tuttavia essi hanno moltissima influenza sui processi di regolazione della cellula. Il meccanismo con cui gli ormoni esercitano il controllo è facilmente spiegato dall'ipotesi della proteina flessibile. Dal momento che queste molecole non sono consumate, esse possono essere usate ripetutamente per attivare le molecole enzimatiche, diversamente dai substrati che, invece, sono consumati. Regolatori di questo tipo sono, dunque, sufficienti in piccola quantità.

In certi casi, come in quello dell'adrenalina, l'ormone iniziale porta alla formazione di una seconda molecola, l'AMP ciclico, che agisce come regolatore per molti enzimi modificandone la forma (si veda l'articolo *AMP ciclico* di Ira Pastan in «Le Scienze», n. 51, novembre 1972). Generalmente le modificazioni di forma indotte dalle molecole regolatrici sono simili a quelle indotte dal substrato.

La regolazione delle vie metaboliche

Nell'esaminare la forma delle proteine è difficile, noioso e spesso inutile indicare dettagliatamente tutte le componenti di una grande molecola proteica. Per comodità, le diverse conformazioni molecolari possono essere rappresentate da figure geometriche: quadrati, triangoli, circonferenze e cosí via (*si veda l'illustrazione a pagina 38*). Ciononostante è importante ricordare che, se anche forme di questo genere possono essere rappresentate mediante semplici figure piane, il passaggio da una circonferenza a un quadrato, per

esempio, sottintende sottili e complessi cambiamenti di orientamento da parte delle numerose catene laterali di amminoacidi che costituiscono una proteina. Questo procedimento è analogo a quello in base al quale si usa il simbolo C per l'atomo di carbonio anziché dare la completa descrizione meccanico-quantistica degli elettroni, protoni e neutroni che compongono l'atomo stesso. Le differenti forme delle proteine in biochimica vengono indicate col nome di conformazioni; esse infatti corrispondono alla definizione chimica di cambiamento di conformazione, cioè alla variazione di forma di una molecola causata dalla rotazione attorno a un singolo legame chimico. I termini conformazione e modificazione di conformazione sono dunque sinonimi tecnici dei termini forma e cambiamento di forma finora usati in questo articolo.

Il punto della superficie enzimatica dove ha luogo l'azione catalitica è chiamato sito attivo. La molecola regolatrice, invece, si lega all'enzima in un punto chiamato sito regolatore o sito allosterico per distinguerlo dal sito catalitico (il termine allosterico, che significa altro sito, fu coniato dal biochimico francese Jacques Monod, un pioniere in queste ricerche, ed è gradualmente entrato nell'uso comune come termine generico per definire le proteine regolatrici).

Le molecole regolatrici vengono anche chiamate effettori, modificatori o effettori allosterici.

Consideriamo ora come queste idee sulla conformazione delle proteine aiutino a spiegare alcuni tipi di regolazione. Uno dei compiti piú importanti di un sistema vivente è quello di stabilire in che modo trattare le sostanze contenute nel cibo. Noi abbiamo bisogno di energia sotto forma di molecole di adenosintrifosfato, o ATP, per tutti i processi dell'organismo: per contrarre i muscoli, per vedere, per attivare i nervi, per rigenerare i tessuti. Questa energia proviene dalla combustione del cibo. Una certa quantità di ATP è costantemente adoperata dall'organismo, anche in stato di riposo; il cuore, per esempio, continua a battere anche quando dormiamo.

Naturalmente è auspicabile che durante i periodi di scarsa domanda di energia si faccia rifornimento in vista delle necessità future. Uno dei modi per rifornirsi di energia consiste nell'accumulare glicogeno, una catena di molecole di zucchero. Quando una molecola di glucosio è ingerita da un sistema vivente essa può prendere due vie: può venire immediatamente ossidata per formare anidride carbonica, liberando grandi quantità di ATP, oppure può venire immagazzinata sotto forma di glicogeno per essere utilizzata a una successiva richiesta (*si veda l'illustrazione a pagina 36*). L'enzima fosfofruttochinasi, che esercita la sua azione lungo la prima di queste due vie, è attivato o inibito dal livello di ATP nel sistema. Un altro enzima, la glicogeno-sintetasi che interviene lungo la seconda via del meccanismo di regolazione del processo di sintesi del glicogeno, viene attivato o inibito dalla presenza di composti regolatori nella cellula. Un terzo enzima, la glicogeno-fosforilasi, catalizza la reazione dal glicogeno al glucosio-1-fosfato lungo una terza via metabolica, ed entra in azione quando si avverte la necessità del recupero della sostanza immagazzinata.

Il funzionamento di questi enzimi, condizionato dalla domanda e dalle scorte, è semplice e ingegnoso. Quando i livelli dell'ATP sono alti, l'enzima fosfofruttochinasi della prima via e l'enzima fosforilasi della terza via sono resi inattivi, mentre viene attivata la sintesi del glicogeno della seconda via. Il glucosio viene dunque avviato all'immagazzinamento sotto forma di glicogeno.

Quando un animale è spaventato, invece, viene secreta adrenalina che a sua volta attiva la fosforilasi e la fosfofruttochinasi, mentre la sintesi del glicogeno viene inibita. In questo caso il glicogeno è rimosso dal deposito e trasformato in ATP che fornisce l'energia necessaria per aiutare l'animale a sfuggire il pericolo. In altre parole, quando siamo spaventati, l'ormone fa sí che le nostre riserve di glicogeno siano rese disponibili come sorgenti di energia. Molti tipi di domanda sono in grado di liberare il glicogeno dal deposito. La fame e alcuni stimolatori dell'attività muscolare (come gli ioni calcio) attivano la fosforilasi e possono cosí generare energia.

In linea di principio le influenze di questo tipo agiscono allo stesso modo dell'adrenalina.

Questo particolare sistema regolatore può essere utile per illustrare due punti importanti. Anzitutto gli enzimi devono essere disponibili in ogni sistema vivente e la loro attività non deve essere costante. La modificazione della conformazione sotto l'influenza di metaboliti e ormoni, dunque, costituisce un meccanismo per rendere attivi o inattivi questi enzimi a seconda del variare delle condizioni esterne (per esempio dalla fame alla sazietà).

Il secondo principio è piú sottile. Se tutti gli enzimi coinvolti nel metabolismo dei carboidrati fossero attivati simultaneamente, il risultato sarebbe solo un corto circuito sulle vie 1, 2 e 3 e il glucosio verrebbe continuamente bruciato e immagazzinato senza alcun vantaggio.

Si è generalmente riscontrato che quando esistono vie alterne di questo tipo una molecola, attivandone una, inibisce l'altra. Nel caso, per esempio, dell'immagazzinamento di glicogeno, la molecola regolatrice inibisce l'enzima della via 2, mentre attiva quelli delle vie 1 e 3. Le vie per l'ossidazione del glucosio e la sua rimozione dal deposito sono accelerate mentre quella dell'immagazzinamento è bloccata. In assenza di questa molecola regolatrice l'equilibrio è favorevole alla forma inattiva dell'enzima nelle vie 1 e 3 e alla forma attiva nella via 2; in altre parole il sistema favorisce l'immagazzinamento del glucosio. In tal modo una modificazione alterna della conformazione degli enzimi impedisce l'instau-

Molecole con funzione di attivatore possono, secondo la teoria della conformazione flessibile, aiutare a far agire da substrato, senza venir consumate, una molecola di per sé inadeguata alterando la conformazione dell'enzima. Per esempio, nel caso di una molecola che sia troppo piccola per indurre il corretto allineamento dei gruppi catalitici (*a*), una seconda molecola può legarsi all'interno del sito attivo in posizione immediatamente adiacente alla molecola insufficiente, e come risultato si ottiene una conformazione stabile che presenta il corretto allineamento dei gruppi catalitici (*b*). Una molecola che si leghi in un sito, al di fuori del sito attivo, può indurre anch'essa una conformazione appropriata (*c*).

rarsi di un corto circuito e assicura che la sintesi del glicogeno venga favorita quando i livelli di ATP sono alti, e che la degradazione ad anidride carbonica venga favorita quando si ha bisogno di energia.

Struttura delle proteine

A questo punto vale la pena di osservare un po' piú dettagliatamente in che modo la conformazione delle proteine effettua questa regolazione. Le proteine che agiscono da enzimi hanno un peso che oscilla tra 10 000 dalton e molti milioni di dalton (1 dalton corrisponde approssimativamente al peso di un atomo di idrogeno).

La cifra piú elevata, tuttavia, è ingannevole perché tutte le proteine di notevoli dimensioni sono costituite da subunità peptidiche che in genere variano da 15 000 a 100 000 dalton. Per esempio l'enzima fosforilasi, che è importante per l'immagazzinamento e la degradazione del glicogeno, è un dimero (polimero di due peptidi) composto da due subunità identiche ciascuna delle quali ha un peso molecolare di 96 000 dalton. Analogamente l'aspartiltranscarbamilasi, il primo enzima sulla via che porta alla sintesi del citidintrifosfato, CTP, è un dodecamero (polimero di dodici peptidi) composto da sei subunità di un tipo (ciascuno con un peso molecolare di 35 000 dalton) e sei subunità di un altro tipo (ciascuno con un peso molecolare di 17 000 dalton). Queste subunità sono attratte fra loro a opera di forze non covalenti: si tratta in larga misura di attrazioni elettrostatiche o legami idrofobi (nome usato per descrivere la forza che spinge le regioni oleose di una struttura ad unirsi insieme nello stesso modo in cui le gocce d'olio in acqua tendono alla coalescenza). Per fortuna le caratteristiche qualitative non cambiano con le dimensioni della proteina, sicché lo studio delle proteine piú semplici permette di capire le proprietà delle proteine in generale.

Una catena peptidica avente un peso molecolare di soli 25 000 dalton è ancora grande, se confrontata con il peso molecolare della maggioranza dei substrati che sono in genere composti il cui peso molecolare oscilla da 100 a 1000 dalton. Occasionalmente gli enzimi agiscono su molecole molto grandi come il DNA (acido desossiribonucleico), la cellulosa o altre proteine, ma in questi casi essi in genere si legano solo con una piccola porzione di queste grosse molecole, di modo che la dimensione effettiva del substrato è ancora di circa 1000 dalton o anche meno. Questa differenza nelle rispettive dimensioni significa che solo una piccola parte della superficie dell'enzima è realmente implicata nella catalisi. Il resto della superficie è disponibile per legare le molecole regolatrici e per l'associazione delle subunità fra loro.

Effettivo cooperativo

Il concetto di flessibilità delle proteine ha offerto una possibilità di spiegazione a un fenomeno noto da tempo e che era stato originariamente scoperto dal fisiologo danese Christian Bohr nel diciannovesimo secolo. Bohr osservò che la curva di affinità dell'emoglobina per l'ossigeno aveva un andamento sigmoide, ed era pertanto diversa dalla normale curva iperbolica che si osserva nella maggior parte degli enzimi (*si veda l'illustrazione a pagina 37*). Egli ne dedusse, a ragione, che questo insolito tipo di curva si sarebbe avuta qualora il primo legame molecolare avesse reso piú facile la formazione del legame alla molecola successiva, e cosí via; pertanto egli definí questo processo « cooperativo ». Da allora l'emoglobina è stata fatta oggetto di numerosi studi da parte di molti ricercatori e la maggior parte delle nostre conoscenze sul fenomeno dell'effetto cooperativo proviene da ricerche eseguite su questa proteina fondamentale. Il fenomeno, tuttavia, non è limitato all'emoglobina. Le curve sigmoidi sono una caratteristica comune e importante anche delle proteine regolatrici.

La comparsa di una curva di affinità ad andamento sigmoide nel caso di una proteina regolatrice può essere spiegata in base all'ipotesi della proteina flessibile usando la piú semplice proteina a piú subunità, e cioè un dimero costituito da due subunità identiche (*si veda l'illustrazione a pagina 39*). Un certo numero di tali dimeri esiste in natura; i tipi di interazione cui essi partecipano sono simili a quelli delle proteine piú complesse. Il legame con il substrato induce modificazioni di conformazione che dipendono tanto dalla struttura del substrato quanto dalla struttura della proteina.

Sono noti tre tipi generali di modificazione di conformazione. In un tipo la prima molecola di substrato altera la subunità alla quale è legata ma non altera le interazioni fra le subunità. La seconda subunità, pertanto, si lega con il substrato proprio allo stesso modo e con la stessa affinità della prima. Non c'è alcuna collaborazione fra i siti delle subunità.

Il secondo tipo ha caratteristiche assai diverse. Qui la prima molecola che deve essere legata induce una modificazione di conformazione nella prima subunità che a sua volta induce una modificazione nella seconda subunità.

Queste modificazioni, che vengono trasmesse attraverso la struttura della proteina, alterano il sito attivo della seconda subunità in modo tale da renderlo piú recettivo verso il substrato; pertanto la seconda molecola del substrato viene legata piú rapidamente della prima. Questo fenomeno, chiamato effetto cooperativo positivo, spiega la curva sigmoide scoperta da Bohr.

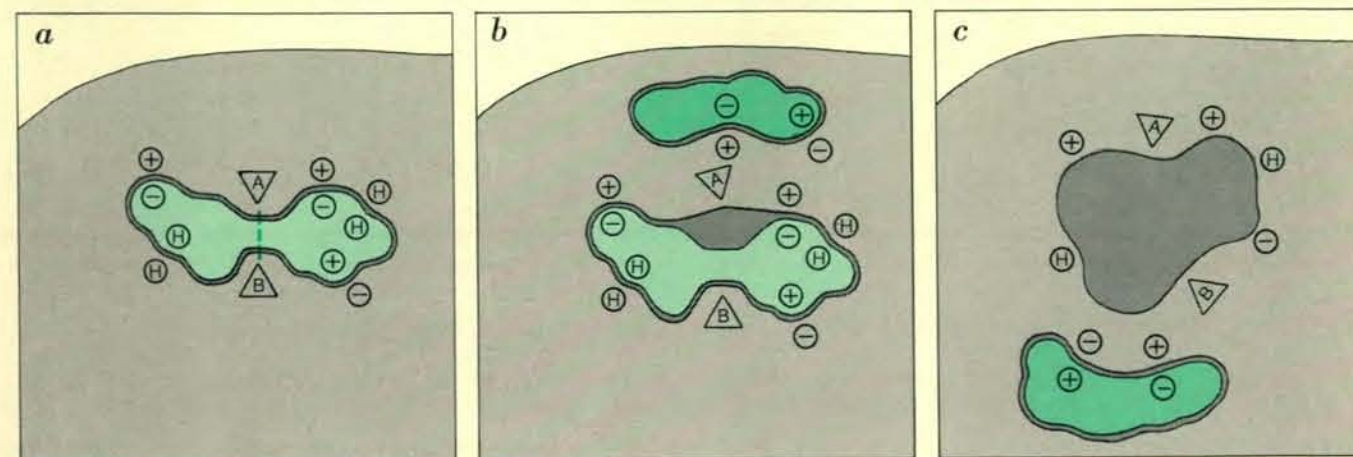
Nel terzo tipo la prima molecola induce una modificazione di conformazione che rende il sito attivo della seconda subunità meno recettivo per il substrato a causa di un'incompatibilità di configurazioni geometriche. Il legame con la seconda molecola diviene dunque piú difficile e si attua molto meno rapidamente. Questo fenomeno viene definito effetto cooperativo negativo (cooperativo in quanto esiste una interazione fra le subunità, negativo in quanto la prima molecola ha un effetto negativo sulla seconda).

La teoria della complementarità indotta può facilmente spiegare la curva sigmoide delle proteine composte da molte subunità; a questo scopo George Nemethy, David Filmer e io abbiamo messo a punto un metodo matematico per verificare tale ipotesi. La soluzione da noi proposta non è la sola spiegazione possibile, ma esistono prove documentate che ne dimostrano la validità per un gran numero di enzimi che presentano curve sigmoidi di questo tipo. Oltre a rendere possibile la spiegazione di un curioso fenomeno in termini di struttura proteica, la teoria della complementarità indotta prevedeva l'esistenza di un diverso tipo di interazione, la cooperazione negativa, dovuta al fatto che le modificazioni di conformazione indotte nella prima molecola rendevano piú difficile il legame con la seconda. Questo fenomeno non era noto, allora, in natura, ma la sua previsione teorica indusse a ricercarlo deliberatamente. Nel 1968 Abby Conway e io ne scoprimmo un esempio in un enzima che partecipa al metabolismo dei carboidrati: la gliceraldeide 3-fosfatodeidrogenasi. L'effetto cooperativo negativo è stato dimostrato in seguito in molti altri enzimi.

L'analisi matematica di tale modello di cooperazione negativa spiega in parte perché il fenomeno era sfuggito all'osservazione. A un esame superficiale una curva con effetto cooperativo negativo appare come un'iperbole. Solo un'analisi piú accurata mostra che in realtà non è tale. Una curva del genere può essere spiegata solo assumendo che la seconda molecola si leghi alla proteina meno rapidamente della prima. Ora sono stati resi noti molti esempi di ciascun tipo di cooperazione. Le modificazioni di conformazioni indotte possono essere causate sia da attivatori e inibitori sia dai substrati, e perciò tutti e tre i tipi di molecola possono dar luogo a interazioni di tipo cooperativo.

Tutti questi concetti si applicano analogamente a molecole aventi piú di due subunità. Nelle interazioni positive la prima molecola facilita il legarsi della seconda, la seconda facilita il legarsi della terza, e cosí via. Nelle interazioni negative ogni molecola che interagisce rende piú difficile l'interazione alla successiva. La piccola molecola (substrato, inibitore o attivatore) induce la modificazione e l'effetto viene quindi trasmesso alle subunità vicine. Quanto piú grande è il numero delle subunità, tanto piú spettacolare può essere la cooperazione, ma il modello che in realtà si osserva dipende dai particolari della struttura individuale della proteina.

Perché dovrebbe esistere in natura un fenomeno di cooperazione di questo tipo? Non sarebbe meglio che la proteina avesse la forma giusta sin dal principio senza che vi fosse la necessità di alterare le interazioni tra le subunità o la forma delle proteine per mezzo di codificazioni di conformazione? Una delle ragioni per cui la flessibilità delle proteine è cosí importante è stata già menzionata: la flessibilità rende possibile la regolazione di un enzima a opera di molecole che non vengono consumate nella reazione enzimatica.

Con il concetto di cooperazione, tuttavia, la flessibilità delle proteine sembra assumere anche un altro significato. Se si esaminano le curve di affinità corrispondenti ai tre tipi di effetto cooperativo e si confrontano le modificazioni di concentrazione di un composto con il livello di attività, si osserva che occorre all'incirca un aumento di concentrazione del substrato di 81 volte per passare da un livello di attività del 10 per cento a uno del 90 per cento, posto che la proteina presenti la tipica curva iperbolica caratteristica di una normale proteina non interattiva. Ben diverso è il caso se la proteina presenta un modesto grado di cooperazione positiva: basterà che la concentrazione del substrato aumenti di nove volte (nell'emoglobina basterà un aumento di concentrazione di tre volte, e nella CTP-sintetasi un aumento di una volta e mezzo). In altre parole, una proteina con effetto cooperativo positivo è molto piú sensibile a piccole fluttuazioni dell'ambiente di quanto non lo sia una proteina normale. La sensibilità è aumentata sia per gli ini-

Molecole con funzione di inibitore possono, con un processo inverso, distorcere un enzima flessibile facendogli assumere una conformazione diversa da quella attiva contraendo dei legami in sedi diverse dal sito attivo. Nel caso di un substrato che in altre condizioni reagirebbe con l'enzima (*a*), tali inibitori possono alterare o la funzione catalitica (*b*), o la capacità di legarsi (*c*) dell'enzima, e in tal modo la reazione tra l'enzima e il substrato viene praticamente annullata.

bitori e gli attivatori sia per i substrati. La cooperazione positiva è pertanto uno strumento di amplificazione che permette a un piccolo segnale di avere un effetto regolatore molto piú ampio. Questo incremento della sensibilità è straordinariamente importante per la funzione regolatrice di queste proteine.

Un esempio dell'importanza fisiologica del fenomeno della cooperazione può essere osservato in pazienti che hanno un'emoglobina mutante, priva dell'effetto cooperativo positivo presente nell'emoglobina normale. La cooperazione positiva consente alle molecole di emoglobina di assorbire grandi quantità di ossigeno nei polmoni e depositarle nei tessuti anche se la pressione di ossigeno non varia molto da una sede all'altra. Questa capacità è effetto della forte pendenza della curva sigmoide associata all'effetto cooperativo positivo. La proteina non cooperativa è invece molto meno efficiente nel trasportare ossigeno, con grave danno dei portatori. In modo un po' grossolano si potrebbe dire che la proteina dotata di effetto cooperativo positivo è come un autocarro che si carichi completamente di rifiuti in un posto e li scarichi tutti in un altro, mentre una proteina che ne è priva non può fare un pieno carico all'inizio e se ne può liberare solo in parte.

Qual è, allora, il ruolo della cooperazione negativa? Uno sguardo alla curva corrispondente fornisce una risposta anche a questa domanda. In questo caso la proteina è meno sensibile alle fluttuazioni dell'ambiente di quanto non lo sia una proteina non cooperativa; in altre parole, è necessario un aumento molto maggiore della concentrazione del substrato, inibitore o attivatore, perché l'attività enzimatica passi da un livello del 10 per cento a uno del 90 per cento. Alcune proteine non devono essere soggette alle fluttuazioni nell'ambiente e le proteine che regolano questi processi possono avvalersi dell'effetto cooperativo negativo per smorzarne la sensibilità e proteggerle dalle modificazioni ambientali.

Le modificazioni di conformazione

Come si propagano le modificazioni di conformazione attraverso la molecola proteica, e di che entità sono? La conoscenza di questi processi ha preso le mosse da due fonti: gli studi chimici della composizione proteica e l'esame ai raggi X della struttura proteica.

Alcuni importanti dati strutturali sono stati ottenuti dall'opera pionieristica di M.F. Perutz e dei suoi collaboratori sull'emoglobina. L'emoglobina è costituita da quattro subunità simili, ma non identiche, disposte in forma di tetraedro. Quando l'ossigeno è legato al gruppo eme, contenente ferro, di una di queste subunità, si verifica un lieve spostamento spaziale dell'atomo di ferro, che a sua volta determina un piccolo spostamento di una catena laterale adiacente di istidina. Questa esercita una compressione su una serie di catene laterali di amminoacidi nella regione a elica della proteina e provoca la fuoriuscita da una piccola sacca di una catena laterale di tirosina. La subunità a cui questo amminoacido è attaccato viene a sua volta spostata rompendo il legame salino tra le subunità.

Queste non sono le sole regioni colpite dal legame contratto con l'atomo di ossigeno, ma l'esempio serve a illustrare il tipo di modificazione che può verificarsi in una molecola proteica. Il punto importante è che la proteina è una molecola saldamente strutturata con intimi rapporti di aderenza fra gli atomi delle catene laterali di amminoacidi. Uno spostamento di uno di questi gruppi, anche di una frazione di angstrom, si ripercuote su altre catene laterali propagandosi lungo tutta la molecola proteica. Le modificazioni all'interno di una subunità dell'emoglobina determinano una ristrutturazione dei rapporti fra le subunità.

L'importanza di questi spostamenti nella proteina può essere dimostrata con emoglobine mutanti isolate da pazienti malati. Quando Perutz e i suoi collaboratori esaminarono le emoglobine mutanti, osservarono che le catene laterali di amminoacidi interessate nella modificazione di conformazione erano alterate.

Una ulteriore comprensione di questi processi è risultata dagli esperimenti condotti nel mio laboratorio all'Università di California a Berkeley da Alexander Levitzki e William Stallcup. Questi autori hanno lavorato sulla CTP-sintetasi, enzima fondamentale che converte l'uridintrifosfato, o UTP, in CTP nel metabolismo degli acidi nucleici. Questa proteina regolatrice in genere esiste in forma di quattro subunità identiche, ma può essere studiata anche come dimero. Ogni subunità possiede dei siti attivi per i suoi substrati: UTP, ATP, NH_3 e glutammina, e un sito allosterico per il suo effettore regolatore guanosintrifosfato o GTP (*si veda l'illustrazione a pagina 40*). La glutammina si lega al sito glutammile dell'enzima formando quello che viene chiamato il complesso glutammil-enzima. In questa reazione si libera dell'ammoniaca (NH_3) che si lega all'apposito sito dell'enzima. L'ammoniaca, a sua volta, reagisce con l'UTP, che reagisce con l'ATP. Lo stretto accoppiamento delle reazioni chimiche indica che i siti sono immediatamente adiacenti l'uno all'altro.

Nella formazione del complesso glutammil-enzima la reazione produce un legame covalente fra la glutammina e una catena laterale di cisteina sulla superficie della proteina. All'inizio i due gruppi cisteina dei due siti attivi sono egualmente reattivi. Quando la glutammina reagisce con la stessa catena laterale, tuttavia, accade una cosa strana. La reazione del substrato con una subunità blocca la catena laterale di cisteina dell'altra subunità. L'effetto è solo temporaneo; il complesso glutammil-enzima reagisce ancora per rigenerare l'enzima libero.

Questa modificazione di conformazione dell'enzima può essere «congelata» (un po' come si ferma un film a un singolo fotogramma), con l'aiuto di un «marcatore di affinità», la diossiazonorleucina. La molecola di diossiazonorleucina è abbastanza simile alla molecola di glutammina da formare un legame covalente con la catena laterale di cisteina, ma tuttavia troppo diversa per poter reagire ulteriormente. Il risultato è che solo una delle due subunità, inizialmente identiche, reagisce con la diossiazonorleucina. Una reazione a livello di una subunità blocca quella che le è vicina, sicché ogni volta reagisce una sola delle due potenziali subunità, e si verifica il fenomeno che abbiamo definito della «metà dei siti».

Questo fenomeno si osserva in vari enzimi e indica che in molti casi le modificazioni di conformazione possono essere trasmesse a grande distanza nella molecola. In base alle dimensioni delle subunità enzimatiche si può calcolare che le catene laterali di cisteina nella CTP-sintetasi sono probabilmente separate da una distanza di 20-60 angstrom.

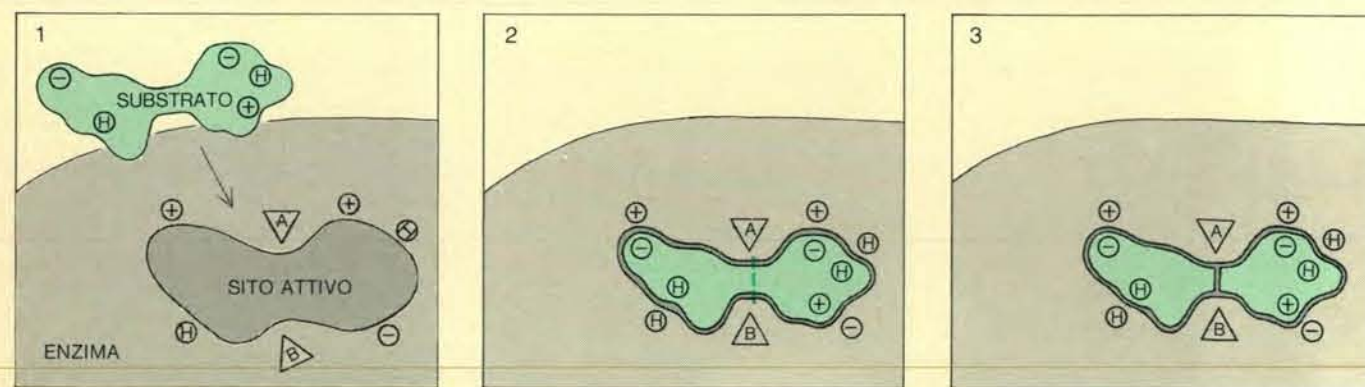
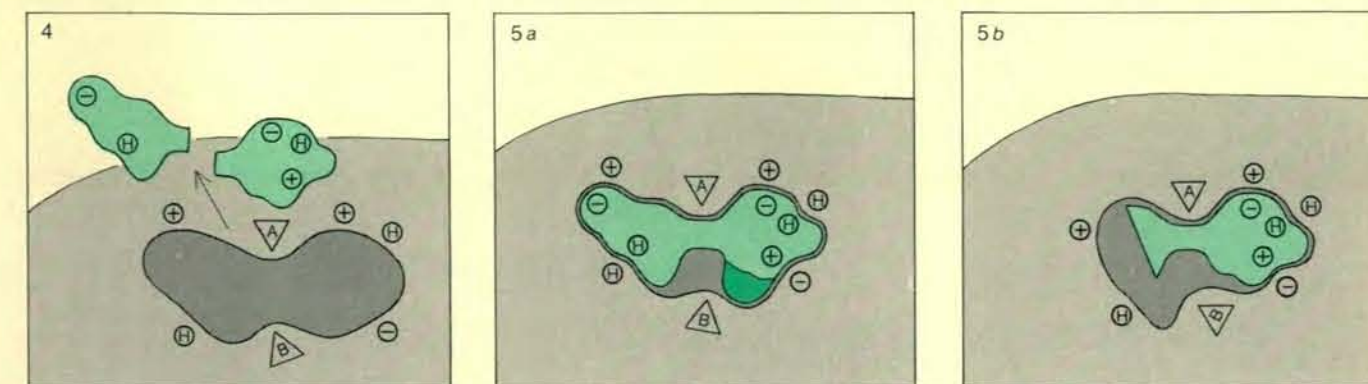
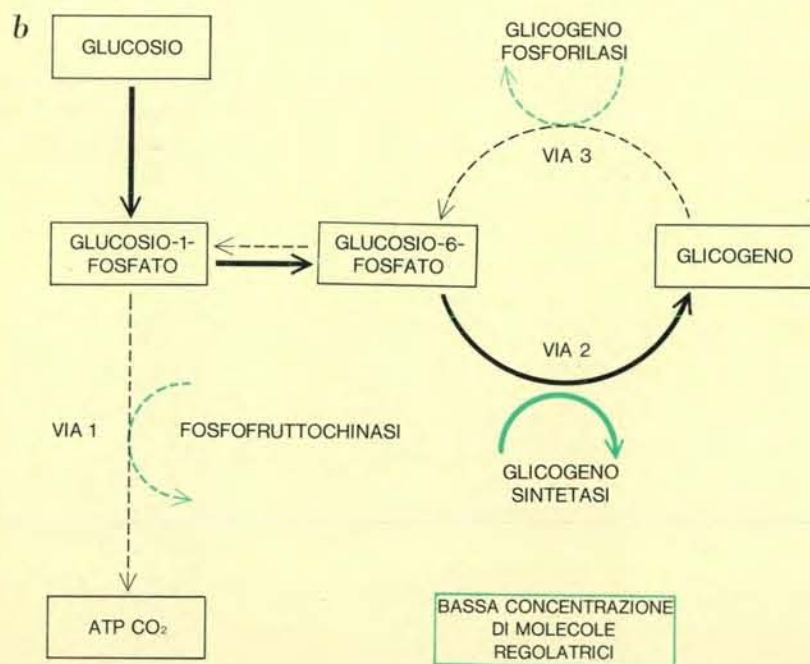
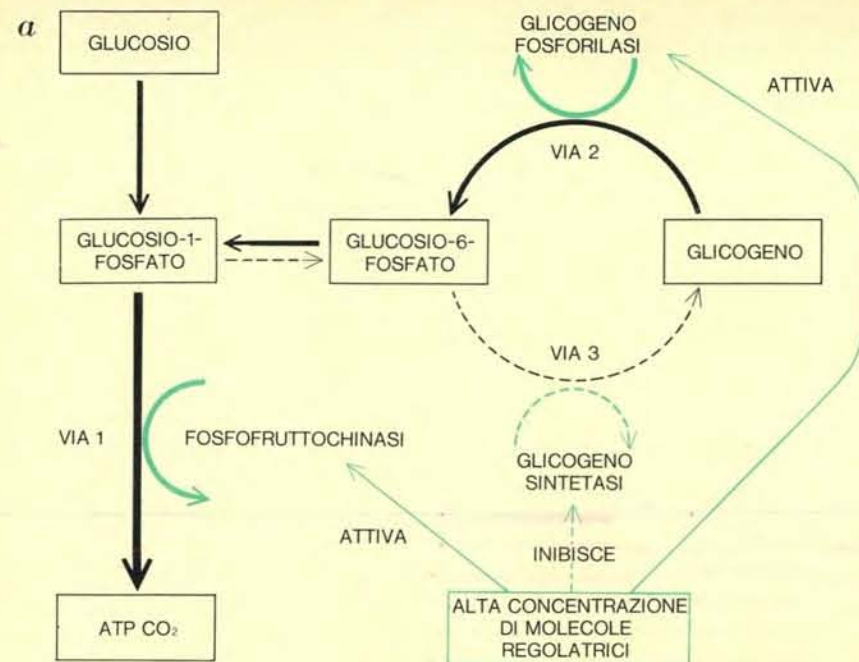
La parte piú sorprendente di questa affermazione è costituita dal fatto che i siti immediatamente adiacenti – ammoniaca, UTP e ATP – non sono alterati affatto dalla modificazione del gruppo cisteina. Tutte le loro proprietà rimangono immutate. Pertanto la formazione di un legame a quanto pare invia a una distanza di 60 angstrom un segnale che ha conseguenze drammatiche, mentre le strutture che si trovano entro il raggio di quattro o cinque angstrom non vengono alterate. Quando si esamina la reattività delle catene laterali di amminoacidi ai reagenti proteici, si osserva che molte catene laterali cambiano posizione e molte altre rimangono immutate.

Un quadro analogo offre il GTP, la molecola regolatrice. Il GTP attiva l'enzima alterando la reattività del sito della glutammina sulla stessa subunità, per cui il complesso glutammil-enzima si forma piú rapidamente. La stessa modificazione di conformazione che attiva il sito glutammile disattiva il sito del GTP della subunità vicina. Per giunta, le modificazioni di conformazione indotte dal GTP non alterano le proprietà dei siti ammoniaca, ATP e UTP. Cosí il GTP, semplicemente legandosi alla superficie della proteina, può provocare l'alterazione di alcune catene laterali lasciando immodificate le altre.

Entità delle modificazioni di conformazione

Sembra dunque che le modificazioni di conformazione indotte nelle proteine non si propaghino come gli anelli concentrici provocati dalla caduta di un sasso in uno stagno. Si tratta piuttosto di qualcosa di simile a una ragnatela i cui fili sono predisposti in modo tale da trasmettere un urto da un'estremità all'estremità opposta della tela. L'urto può essere trasmesso a grande distanza e può alterare la posizione di molti fili, ma un disegno ben fatto può garantire che alcuni fili rimangano immodificati mentre altri si spostano in misura apprezzabile. La proteina, come la ragnatela, è predisposta per trasmettere l'informazione in modo selettivo ad alcune regioni lasciandone altre inalterate.

I disegni schematici presentati in questo articolo per illustrare l'alterazione della conformazione proteica tendono a esagerare i movimenti necessari a ottenere questo controllo. In realtà non sappiamo di che entità deve essere il movimento delle catene laterali per ottenere l'innesco o l'interruzione del meccanismo. Molti elementi, tuttavia, starebbero a indicare come questi movimenti non debbano necessariamente essere molto ampi. La lunghezza di un legame carbonio-ossigeno, ad esempio, è di circa 1,3 angstrom. Un gruppo catalitico che avesse bisogno di stare abbastanza vicino a un atomo di ossigeno per poterne estrarre gli elettroni sarebbe pertanto inefficace come catalizzatore se fosse posto vicino all'atomo di carbonio. Pertanto un movimento dell'ordine di uno o due angstrom dovrebbe essere sufficiente a stabilire tutta la differenza fra un catalizzatore efficiente e uno inefficiente. Il movimento, per essere efficiente, può essere di entità anche minore, tanto piú che la maggior parte dei regolatori non rende completamente inattivo l'enzima, ma piuttosto ne riduce la funzione in direzione o della catalisi o della formazione di legami.

William N. Lipscomb jr. e i suoi colleghi dell'Università di Harvard con la cristallografia ai raggi X hanno osservato movimenti delle catene laterali di 10-12 angstrom in molecole di carbossipeptidasi. Sembra probabile che

Modello dell'azione enzimatica definita come «complementarità indotta». Secondo questo modello, elaborato in larga misura dall'autore e dai suoi colleghi dell'Università di California, a Berkeley, l'enzima non possiederebbe sin dall'inizio una forma complementare a quella del substrato (*1*), ma verrebbe piuttosto ad assumere una conformazione complementare proprio come l'introduzione di una mano va inducendo un cambiamento nella forma di un guanto (*2*). Una volta legato il substrato, i gruppi catalitici dell'enzima si trovano in posizione tale da rompere il legame chimico nella strozzatura del substrato, dando luogo alla formazione dei prodotti della reazione (*3*); questi lasciano la superficie dell'enzima e il sito attivo ritorna alla sua originaria forma non complementare (*4*). Questo concetto della flessibilità dell'enzima ha permesso di spiegare il fatto, prima difficilmente comprensibile, che certi composti o troppo grandi o troppo piccoli potevano legarsi alla superficie dell'enzima senza dar luogo alla formazione di prodotti (*5a* e *5b*). Sebbene entrambi i composti non-substrato si leghino, nessuno di essi riesce a indurre il giusto allineamento dei gruppi catalitici e pertanto nessuno dei due reagisce con l'enzima.

Molecole aventi effetto regolatore controllano vie biologiche alterne attraverso una modificazione della conformazione degli enzimi. Nel corpo umano, per esempio, il glucosio ingerito viene di norma convertito in glucosio-6-fosfato che può essere o ossidato immediatamente per fornire energia sotto forma di molecole di ATP (adenosintrifosfato) oppure immagazzinato sotto forma di glicogeno (una catena di molecole di zucchero) per essere liberato in futuro in caso di necessità. La via prescelta dipende dalla concentrazione delle molecole regolatrici. Quando vi è fabbisogno di energia, elevate concentrazioni di queste molecole conferiscono all'enzima della via 1 (fosfofruttochinasi) e all'enzima della via 3 (glicogenofosforilasi) la conformazione attiva, mentre al tempo stesso viene conferita una forma inattiva all'enzima della via 2 (glicogenosintetasi), in tal modo incanalando verso la produzione di ATP sia il glucosio ingerito, sia il glicogeno immagazzinato (*in alto*). Basse concentrazioni di molecole regolatrici fanno ritornare alla forma inattiva gli enzimi delle vie 1 e 3 restituendo la sua attività all'enzima della via 2, favorendo così la trasformazione del glucosio in glicogeno con funzioni di deposito (*in basso*). L'effetto reciproco delle molecole regolatrici sulle vie 2 e 3 impedisce la contemporanea attivazione di entrambe.

gli ampi movimenti osservati in questo tipo di proteina siano maggiori di quanto non sarebbe necessario per la regolazione in generale, e che molti degli atomi si muovano molto meno o per nulla. È stato già osservato che i movimenti di alcuni atomi come l'atomo di ferro nell'emoglobina e l'atomo di zolfo nella CTP-sintetasi possono essere inferiori a un angstrom, e tuttavia tale movimento può determinare vistose modificazioni di conformazione. Esperimenti eseguiti con modelli di composti piú semplici fanno ritenere che anche un'alterazione di meno di mezzo angstrom possa modificare in misura notevole la reattività. In breve, i disegni schematici qui mostrati dovrebbero servire a indicare che i movimenti sono abbastanza rilevanti da alterare la funzione di una proteina, ma che essi non rappresentano la misura reale delle modificazioni di conformazione.

Osservazioni generali sulla molecola proteica indicano anche che nella subunità presa nel suo insieme non si verificano modificazioni di forma anche se alcuni atomi si muovono di parecchi angstrom. Se si concentra la propria attenzione su alcuni gruppi e ci si serve di disegni schematici, ci si può formare un'impressione errata, soprattutto se non si tiene presente che si sta guardando solo a una piccolissima porzione della ragnatela. Il grosso dell'anatomia della tela rimane immodificato; solo pochi fili si spostano nelle loro reciproche posizioni, tuttavia questi spostamenti sono altamente significativi in rapporto alla funzione.

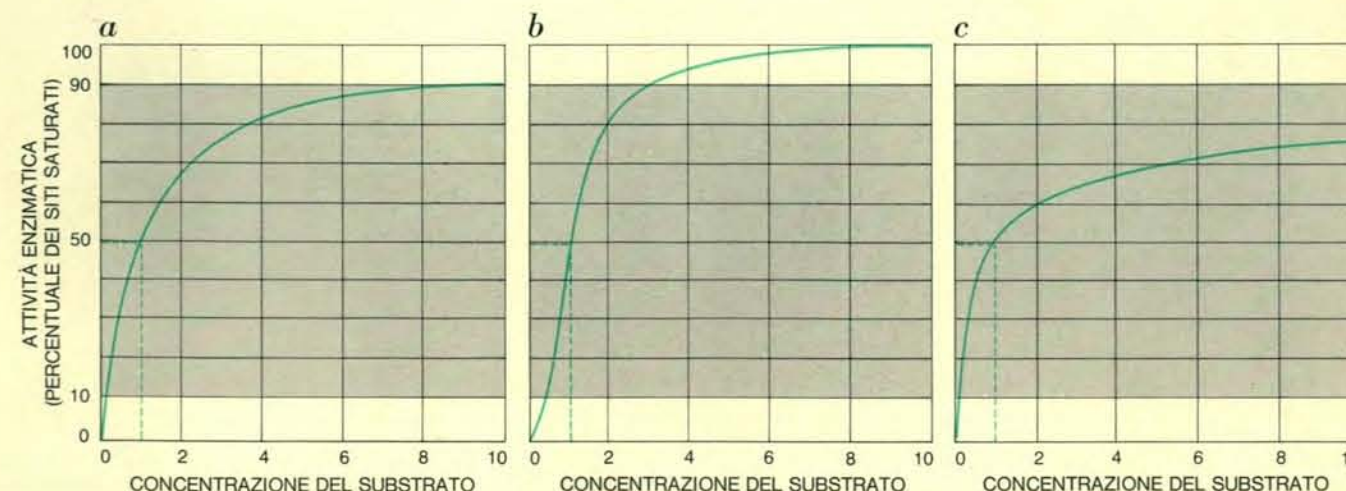
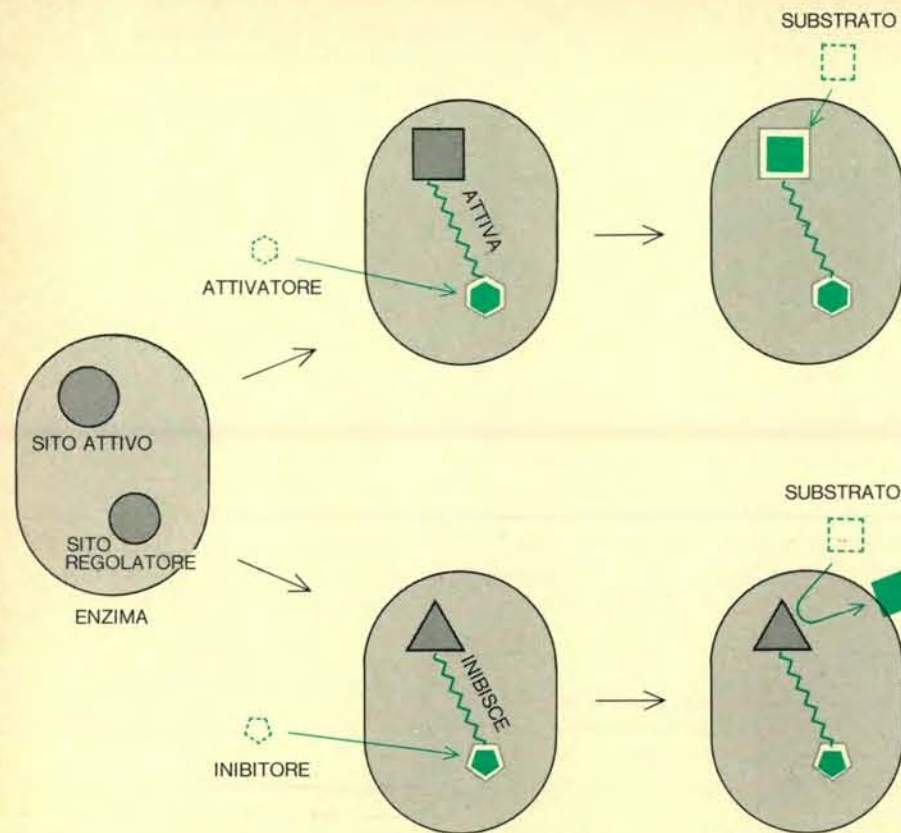
Vantaggi delle modificazioni indotte

Uno dei vantaggi derivanti dal fatto che gli enzimi possono essere indotti a cambiare conformazione in risposta a uno stimolo appropriato è illustrato dall'esempio che abbiamo considerato: la CTP-sintetasi. Ho già detto che i substrati ATP e UTP sono dotati di un forte effetto cooperativo positivo. Nel caso dell'ATP, per esempio, la prima molecola di ATP altera la conformazione dell'enzima in modo tale che le successive molecole di ATP si legano rapidamente. In realtà la CTP-sintetasi è uno degli enzimi piú cooperativi che si conoscano (anche piú dell'emoglobina). Il legame con la prima molecola di substrato ha un effetto innescante che permette il rapido e completo instaurarsi del legame con le successive molecole di ATP, sicché è possibile trovare in quantità apprezzabile solo l'enzima libero e una sua forma con legate quattro molecole di ATP. Questo effetto cooperativo positivo significa che l'enzima è molto suscettibile a piccole fluttuazioni nella concentrazione di ATP ed è perciò altamente sensibile alle influenze ambientali su questa molecola. La molecola regolatrice GTP, invece, induce modificazioni di conformazione che portano a un effetto cooperativo negativo che desensibilizza l'enzima verso fluttuazioni del GTP nell'ambiente.

Cosa significa tutto questo in termini di azione enzimatica? Significa che l'enzima nel corso dell'evoluzione è stato progressivamente modellato al fine di avere una elevata sensibilità nei confronti di variazioni anche modeste dell'ATP nell'ambiente, e una scarsa sensibilità nei confronti di fluttuazioni anche vistose dei livelli del GTP. Questa necessità sorge probabilmente dal fatto che l'ATP, grazie alla sua fondamentale importanza per tante vie metaboliche, è altamente controllato nel sistema biologico. Come risultato, enzimi quali la CTP-sintetasi devono rispondere prontamente a fluttuazioni anche modeste nei livelli di ATP quando sono regolati dal livello dello stesso ATP. D'altra parte, il GTP può fluttuare largamente da una fase all'altra del ciclo vitale dell'organismo perché la sua funzione si esercita in diverse vie, ma meno essenziali. Se è necessario un livello abbastanza costante di CTP-sintetasi attraverso tutte queste fluttuazioni, l'effetto cooperativo negativo del GTP garantirà una scarsa sensibilità verso fluttuazioni nei livelli del GTP. Naturalmente, una desensibilizzazione completa permetterebbe la produzione di CTP in assenza di GTP, operazione inutile, dato che entrambi sono necessari alla sintesi dell'RNA. La proteina assicura il controllo del GTP facendolo agire da attivatore, ma elimina l'eccessiva sensibilità per mezzo dell'effetto cooperativo negativo.

Un altro vantaggio di queste modificazioni di conformazione indotte scaturisce dalla sequenza dei vari stadi di una reazione sulla superficie enzimatica. Nelle sintesi chimiche si formano spesso composti intermedi altamente energetici, ma se essi non sono isolati dall'acqua o da altre sostanze reattive si scompongono in reazioni collaterali che abbassano il rendimento della reazione. Modificazioni di conformazione indotte possono far sí che uno stadio della reazione inneschi lo stadio successivo, che a sua volta innesca il terzo e cosí via. In tal modo i composti intermedi altamente energetici esistono solo per intervalli brevi, e durante le modificazioni chimiche si raccolgono al sicuro nel sito attivo. Sono impedite inutili reazioni collaterali e viene assicurato il rendimento elevato caratteristico delle reazioni enzimatiche. In breve, il cambiamento di conformazione che induce la complementarità spiega l'anomala necessità di una precisa sequenza nella formazione dei legami con i substrati e al tempo stesso ne fornisce una ragione.

È probabile, dunque, che gli enzimi flessibili si siano sviluppati precocemente nel corso dell'evoluzione sotto la spinta delle necessità catalitiche e di certi fabbisogni specifici. Tra i principali vi era quello dell'esclusione dell'acqua da alcune reazioni. La proprietà della cooperazione probabilmente si sviluppò piú tardi in conseguenza della struttura a molte subunità dell'enzima; la cooperazione sopravvisse perché aveva una propria utilità: l'amplificazione di alcune risposte e l'attenuazione di altre.

È importante notare che il parallelismo tra attivatori e inibitori (effettori positivi ed effettori negativi) da un lato, e effetto cooperativo positivo e negativo dall'altro non dovrebbe oscurare il fatto che le due classi di fenomeni hanno ruoli differenti e complementari. La funzione di un attivatore è di dare inizio a una reazione enzimatica e quella di un inibitore di interromperla. L'effetto cooperativo serve ad aumentare o a diminuire la sensibilità dell'enzima alle fluttuazioni ambientali di questi regolatori. Un attivatore può essere cooperativo in senso positivo o negativo, o può non avere nessun effetto cooperativo, e la stessa cosa vale

Le curve qui riprodotte illustrano tre tipi di effetto cooperativo osservati in diverse combinazioni enzima-substrato. Le curve sono ottenute mettendo in rapporto diverse combinazioni di substrato con il livello di attività enzimatica indotta dal substrato. La curva A mostra una tipica curva iperbolica quale si osserva nella maggior parte degli enzimi; è necessario un aumento di 81 volte nella concentrazione del substrato per ottenere un incremento dell'attività dal 10 al 90 per cento. La curva B è una curva ad andamento sigmoide tipica di un enzima con effetto cooperativo positivo (la prima molecola legata facilita il legame con la seconda e cosí via); in questo caso è necessario che la concentrazione del substrato aumenti di sole nove volte perché il livello di attività passi dal 10 al 90 per cento. La curva C è il tipo di curva caratteristico di un enzima con effetto cooperativo negativo (la prima molecola legata rende piú difficile il legame con la seconda molecola e cosí via); la curva è abbastanza simile a un'iperbole, ma in realtà non lo è. La curva si avvicina cosí lentamente allo stato di saturazione finale da non raggiungere nemmeno il livello di attività del 90 per cento. In realtà sarebbe necessario un aumento della concentrazione del substrato di 6541 volte perché un enzima di questo tipo passi dal 10 al 90 per cento di attività. Per rendere evidente il rapporto tra i vari tipi di effetto cooperativo le tre curve sono disposte in modo tale che a una concentrazione del substrato di 1 corrisponda una situazione in cui la metà dei siti è occupata. Curve simili possono essere ottenute sia per molecole con funzioni di attivatore sia per molecole con funzioni di inibitore.

Schema semplificato che illustra il meccanismo con cui l'attività enzimatica viene controllata attraverso modificazioni di conformazione attuate da piccole molecole regolatrici che rimangono estranee alla reazione enzimatica fondamentale. La catena polipeptidica complessa che costituisce l'enzima vero e proprio è rappresentata dall'area in grigio chiaro; le figure in grigio scuro rappresentano simbolicamente le diverse conformazioni del sito attivo e del sito allosterico dell'enzima in assenza di qualsiasi legame con altre molecole. Se viene legata una molecola con funzione di attivatore (*esagono colorato*), questa induce una modificazione di conformazione nel sito allosterico (*espresso simbolicamente come un passaggio da un circolo a un esagono*), e inoltre altera anche la struttura dell'enzima (*linea spezzata colorata*) che a sua volta modifica la conformazione del sito attivo (*passaggio da un circolo a un quadrato*). Come risultato il substrato (*quadrato in colore*) può ora legarsi più facilmente che non in assenza dell'attivatore, in quanto non ha bisogno, per legarsi, di alcuna ulteriore modificazione di conformazione. Nel processo inverso la molecola con funzione di inibitore (*pentagono in colore*) induce un diverso tipo di modificazione di conformazione nel sito allosterico (*passaggio da un circolo a un pentagono*) che viene trasmessa attraverso la struttura dell'enzima col risultato che l'alterazione conseguente nel sito attivo (*passaggio da un circolo a un triangolo*) fa sí che il substrato venga respinto.

per un substrato o un inibitore. Il GTP, per esempio, è un attivatore dell'azione enzimatica, ma la sua curva di affinità mostra un effetto cooperativo negativo. Per esprimersi con semplicità si potrebbe dire che l'attivazione e l'inibizione definiscono il ruolo fondamentale delle molecole regolatrici, mentre l'effetto cooperativo fornisce la fine sintonia del sistema.

Finalmente occorre far osservare che, sebbene nella nostra esposizione della teoria della complementarità indotta ci siamo soffermati sulla particolare impostazione teorica adottata nel nostro laboratorio, molti altri studiosi, sia negli Stati Uniti, che in Europa hanno fornito contributi di grande rilievo alla nostra comprensione del ruolo svolto dalla conformazione delle proteine nel controllo biologico.

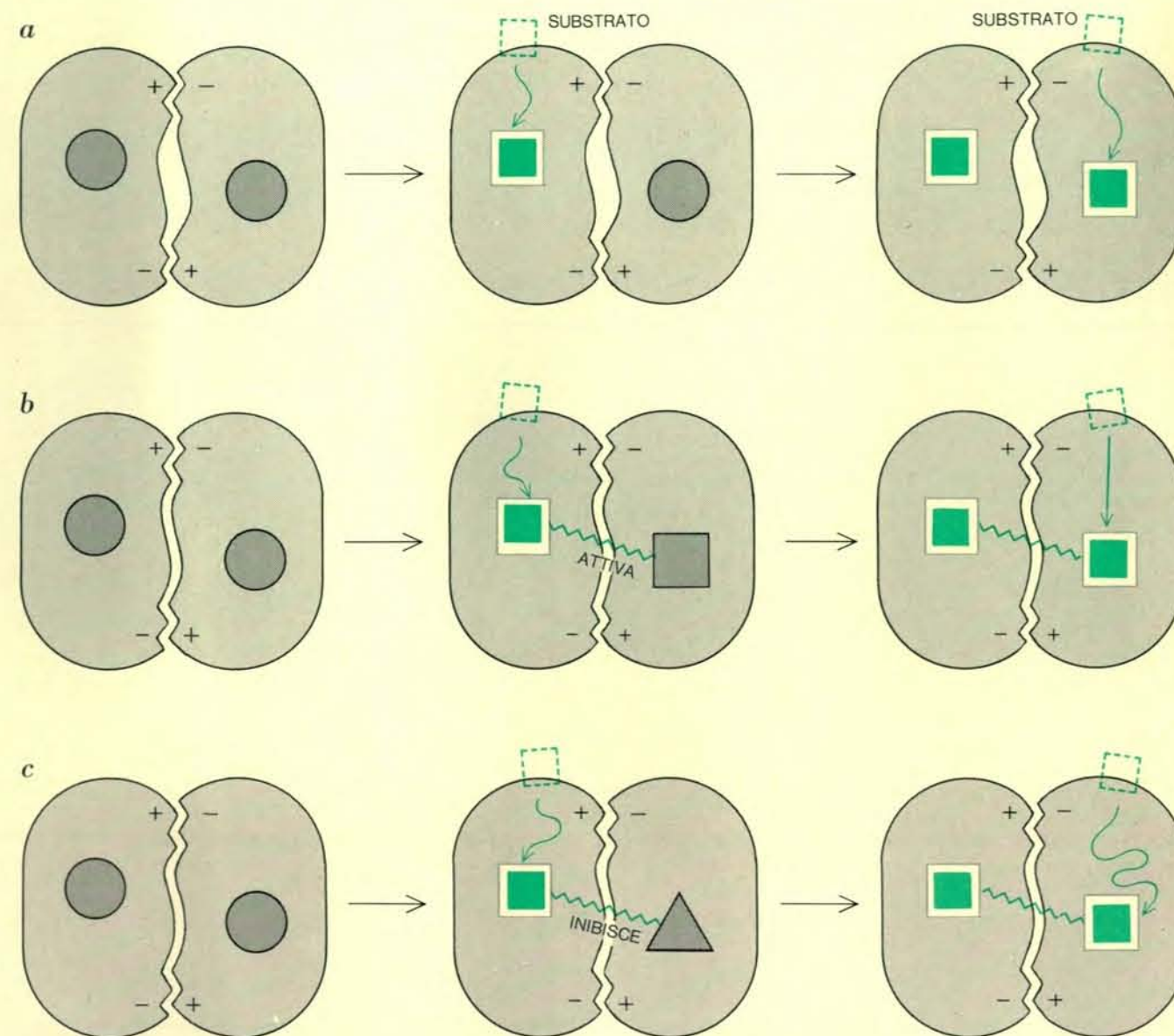
Inoltre, sono state avanzate diverse altre possibili spiegazioni di alcune delle proprietà regolatrici delle proteine.

In particolare Monod, Jeffries Wyman e Jean Pierre Changeux dell'Istituto Pasteur hanno suggerito una spiegazione dell'effetto cooperativo con caratteristiche molto diverse dal meccanismo della complementarità indotta illustrato qui. Può essere dimostrato matematicamente che il modello da loro proposto può spiegare l'effetto cooperativo positivo e anche molte caratteristiche dell'attivazione e dell'inibizione a opera di molecole effettrici. Molti elementi del loro modello sembrano essere presenti nel legame cooperativo fra emoglobina e ossigeno e nelle proprietà di taluni enzimi; inoltre, esso è stato di grande aiuto per chiarire alcune proprietà delle proteine regolatrici. La loro teoria non può però applicarsi a tutti gli enzimi perché l'effetto cooperativo negativo può essere spiegato solo con l'ipotesi della complementarità indotta. Inoltre, Sidney A. Bernhard e i suoi collaboratori dell'Università dell'Oregon hanno postulato che, in alcuni casi, un'asimmetria del tipo di quella osservata nei cristalli di insulina da Dorothy Crowfoot Hodgkin e dai suoi colleghi della Università di Oxford possa avere un ruolo importante nelle proteine regolatrici. Nonostante l'utilità di queste alternative nella descrizione di alcune proprietà delle proteine, sembrerebbe che il meccanismo fondamentale in natura per il controllo biologico sia l'alterazione indotta della struttura proteica così come è stata qui descritta.

L'innesco della modificazione

Ho finora menzionato due tipi di eventi che possono innescare modificazioni di conformazione: 1) la reazione di un substrato per formare un nuovo legame covalente con una catena laterale di amminoacidi, e 2) il legarsi di una molecola regolatrice (substrato, attivatore o inibitore) alla superficie di una proteina senza formazione di un legame covalente. Entrambi questi eventi si verificano nella reazione dei substrati e nei processi regolatori. Negli esempi precedenti mi sono servito solo del tipo di legame contratto da effettori quali il GTP per illustrare la regolazione, ma nei siti allosterici può verificarsi anche una regolazione mediante formazione di legami covalenti. La fosforilazione dell'enzima fosforilasi a opera dell'ATP, per esempio, può bloccare l'enzima in una nuova struttura avente diverse proprietà di reattività e di sensibilità nei confronti dei composti regolatori. Earl W. Sutherland jr. e i suoi colleghi della facoltà di medicina della Vanderbilt University hanno dimostrato che questa formazione di un legame covalente è controllata dall'ormone adrenalina attraverso il suo messaggero, l'AMP ciclico. Analoghe modificazioni sono causate dall'adenilazione dell'enzima glutamminsintetasi, come è stato dimostrato da Earl R. Stadtman, Helmut Holzer e i loro collaboratori al National Heart and Lung Institute. Questi legami covalenti possono essere usati per attivare o inibire l'enzima. In questi casi il ritorno all'enzima originale è ottenuto per mezzo di un secondo enzima che rompe il legame covalente della molecola regolatrice.

Perché la natura dovrebbe servirsi di due mezzi, la formazione di legami covalenti e la formazione di legami non covalenti, per indurre lo stesso tipo di modificazione di conformazione? Per ottenere la regolazione? La risposta non ci è nota con certezza, ma sorge il sospetto che uno dei fattori in gioco sia il tempo. Un animale spaventato e che abbia bisogno di grandi quantità di energia per un breve periodo di tempo dovrebbe poter avere un meccanismo di controllo che assicuri l'elevata attività di alcuni enzimi fondamentali finché la crisi non sia passata. La fosforilazione della fosforilasi indotta dall'adrenalina sembrerebbe fornire un meccanismo del genere. La formazione di un legame covalente fosforil-enzima modifica la conformazione dell'enzima rendendolo piú attivo. Quando la crisi è passata, il gruppo fosfato può essere rimosso per mezzo di un secondo processo enzimatico per rigenerare gli enzimi originali, più torpidi, ma per un certo tempo i normali controlli istantanei sono stati eliminati allo scopo di mobilizzare il glucosio necessario per la crisi.

Analogamente, le modificazioni di conformazione che si verificano in seguito all'aggregazione di due o più subunità possono svolgere anch'esse un ruolo importante nella regolazione. In molti casi si osserva che una proteina è attiva come monomero ma inattiva come dimero, oppure inattiva come dimero e attiva come tetramero. L'associazione delle catene polipeptidiche tra loro determina un cambiamento di conformazione che dà come risultato un'alterazione dell'attività della proteina. In sostanza, dunque, una subunità diventa il regolatore dell'altra. Se questo è vero, i composti che causano direttamente modificazioni di conformazione dovrebbero anche indurre cambiamenti indiretti nella struttura polimerica delle proteine. In realtà, alcuni composti inducono modificazioni di forma che ne causano la dissociazione. Queste reazioni di associazione-dissociazione sembrano essere importanti per molti enzimi nelle vie metaboliche. Le modificazioni di conformazione che si verificano in queste associazioni e dissociazioni sono esattamente analoghe a quelle indotte dalle piccole molecole. È probabile anche che certi ormoni agiscano come regolatori proprio allo stesso modo in cui

In questi diagrammi l'effetto cooperativo è spiegato sulla base del modello dell'attività enzimatica definito della « complementarità indotta ». Il tipo qui rappresentato è il più semplice enzima a subunità che si conosca: un dimero (cioè un polimero formato da due peptidi) costituito da due subunità identiche. I tre diversi enzimi illustrati sono rispettivamente non cooperativi (*fila in alto*), hanno un effetto cooperativo positivo (*fila in mezzo*), e un effetto cooperativo negativo (*fila in basso*).

agiscono in altri casi piccole molecole o subunità.

I cambiamenti di conformazione nelle proteine possono verificarsi molto rapidamente, in alcuni casi solo in un bilionesimo di secondo. Altri si verificano nell'ambito di un millisecondo, e si è dimostrato che questi limitano la velocità di alcune reazioni enzimatiche; altri ancora hanno bisogno di minuti per verificarsi, e questi per la maggior parte sono implicati nelle reazioni di associazione-dissociazione. Questa varietà di tempi serve bene agli scopi della regolazione perché alcuni processi richiedono una correzione istantanea in risposta a uno stimolo, mentre per impedire una reazione eccessiva o per bloccare una risposta finché la crisi non sia passata è necessaria una risposta piú lenta. Inoltre, è stato recentemente suggerito che un lento cambiamento di conformazione sia alla base di un certo tipo di memoria propria della chemiotassia batterica, cioè quel fenomeno per cui un batterio posto in una soluzione di determinate sostanze chimiche si muove in risposta a un gradiente di concentrazione. Un cambiamento del genere può avere la sua controparte nel sistema nervoso degli animali. Le modificazioni di conformazione possono verificarsi in tempi molto variabili, da molto corti a molto lunghi, e ognuno di questi può offrire i suoi vantaggi.

Applicazione ad altri sistemi

Finora ho messo l'accento sul meccanismo regolatore di controllo, degli enzimi, perché questi sono le proteine regolatrici piú estesamente studiate e perché ve n'è ampia disponibilità in laboratorio. Man mano che la biochimica ha progredito, tuttavia, sono state isolate altre molecole che hanno una funzione fondamentale di regolazione nei sistemi biologici. Si è dimostrato che le molecole aventi funzione di recettore nei sistemi sensoriali sono simili agli enzimi per quanto riguarda la struttura e la proprietà di legarsi. Queste molecole hanno la specificità caratteristica degli enzimi, e si ritiene in genere, pur mancandone la prova assoluta, che le modificazioni di conformazioni indotte siano i segnali che innescano l'impulso sensoriale. Quando noi sentiamo l'odore o il gusto di una sostanza, la molecola recettrice subisce un cambiamento di forma che dà il via a una risposta nel nostro sistema nervoso. La luce modifica la conformazione di una proteina presente nel nostro occhio. I fenomeni sensoriali sono la forma piú alta di regolazione e il cervello è il supremo apparato regolatore del piú complesso sistema biologico: l'uomo.

Analogamente, sono state isolate molecole che controllano la sintesi proteica. Molecole aventi funzione di repressore, per esempio, si legano al DNA e impediscono la lettura del messaggio genetico a meno che non vengano rimosse da induttori. Inoltre si ammette, pur mancando anche in questo caso la prova assoluta, che l'induttore provochi un cambiamento di forma che stacca il repressore dal DNA, dando cosí inizio al processo della sintesi proteica. Si ritiene anche che le modificazioni di conformazione abbiano un'importanza fondamentale nelle molecole con funzioni di promotore, come la proteina ciclica che lega l'AMP che si lega al DNA e contribuisce a dar inizio alla sintesi proteica. Un inizio di questo tipo e il controllo della lettura di porzioni selezionate del DNA permettono di produrre proteine diverse a seconda delle circostanze. Si è ipotizzato che un controllo di questo tipo sia in realtà un meccanismo di differenziazione, il motivo per cui le cellule nervose hanno certe proteine e le cellule muscolari ne hanno altre. La differenziazione, il processo fondamentale che permette all'organismo multicellulare di avere funzioni specializzate, è perciò anch'esso dipendente da modificazioni di conformazione delle proteine.

Le molecole di anticorpo, che ci proteggono contro l'invasione da parte di sostanze estranee, inducono una serie di reazioni nel sistema chiamato fissazione del complemento. Questo sistema viene attivato perché distrugga, assimilandole, cellule e proteine dannose. Per giunta la presenza di una proteina estranea, come per esempio la tossina difterica, può indurre le cellule produttrici di anticorpi a moltiplicarsi rapidamente perché possa essere prodotto l'anticorpo per lo specifico agente dannoso. Questa induzione selettiva di certi tipi di comportamento protettivo costituisce un altro meccanismo regolatore dei sistemi viventi, e si ritiene che anch'esso sia legato a modificazioni indotte di conformazione degli anticorpi o dei complessi anticorpo-recettore. Anche il trasporto delle sostanze nutritizie attraverso la membrana cellulare si considera sia attivato da cambiamenti di conformazione indotti nelle proteine cosí come la contrazione muscolare.

Non esiste un singolo processo di regolazione nei sistemi viventi, cosí come non c'è un singolo tipo di controllo in una grande città. I sistemi di controllo nelle città sono mediati dai vari apparati di giustizia, vigilanza, amministrazione e cosí via mentre controlli nei sistemi biologici sono mediati da enzimi, anticorpi, recettori, repressori eccetera. Ognuna di queste molecole proteiche è designata per scopi diversi e perciò ha una propria storia peculiare, col risultato di una enorme varietà. Ciò nonostante, da tutta questa complessità emerge una grande semplicità. La flessibilità delle proteine e le modificazioni di conformazione cui esse possono andare incontro sono gli elementi fondamentali del meccanismo regolatore di controllo. È perché la proteina può esistere in piú forme e queste possono essere alterate dagli agenti esterni, che il sistema vivente può reagire agli stimoli esterni e proteggersi contro le modificazioni ambientali.

La natura, dunque, pone nella cellula vivente i piú potenti catalizzatori conosciuti e li piega ai suoi comandi per mezzo di movimenti impalpabili all'interno della loro struttura, movimenti dell'ordine di angstrom. Queste modificazioni di forma possono essere immaginati, piú o meno, come il cambiamento di forma di un guanto quando vi viene introdotta una mano, ma a un esame piú approfondito esse rassomigliano piuttosto alle delicate risposte di una ragnatela tessuta con squisito equilibrio. La sottile tela che chiamiamo proteina può essere alterata nella sua forma da minuscoli urti, ed è attraverso le ripercussioni di questi urti che le funzioni possono essere messe in moto o bloccate. Come risultato noi abbiamo fame quando abbiamo bisogno di mangiare, e non l'abbiamo piú quando siamo sazi. Possiamo controllare l'uso dell'energia e la crescita dei tessuti specializzati, proteggerci contro sostanze nocive che ci invadono e sviluppare un cervello.

In questi diagrammi schematici che illustrano il comportamento del dimero della CTP sintetasi, enzima fondamentale del metabolismo degli acidi nucleici, è dimostrata la natura altamente specifica delle modificazioni che intervengono nella struttura e nell'azione delle proteine in seguito al loro legame con un substrato selezionato e con molecole regolatrici. Esperimenti, condotti nel laboratorio dell'autore, hanno dimostrato che esistono siti attivi specifici sulla superficie di questo enzima per quattro molecole di substrato e un sito allosterico per una molecola regolatrice (*a*). Quando l'amminoacido glutammina si lega nel sito glutammile su una delle due subunità dell'enzima (*b*), si ha la formazione di un complesso intermedio glutammil-enzima. Le modificazioni di conformazione indotte in questa subunità causano a loro volta modificazioni nella conformazione del sito glutammile sull'altra subunità riducendone la reattività. Poiché solo una subunità di ciascun dimero può partecipare a questa reazione, tale processo viene a volte definito il fenomeno della «metà dei siti». Il legame fra la molecola regolatrice GTP (guanosintrifosfato) e una delle due subunità, induce invece modificazioni di conformazione che fanno reagire piú rapidamente il sito glutammile, mentre al tempo stesso fa assumere al sito allosterico GTP dell'altra subunità una conformazione diversa e meno affine alla seconda molecola GTP (*c*). In questo caso la proteina è programmata in modo tale da permettere al GTP di agire come attivatore di una reazione catalitica e come inibitore di un'ulteriore assunzione di GTP. Il legame con la glutammina e col GTP determina modificazioni di conformazione trascurabili nei siti ATP, UTP o NH_3.

L'evoluzione delle Ande

La geologia delle Ande centrali indica che la storia di questa catena montuosa può essere spiegata attraverso la distruzione di una zolla della crosta terrestre che si immerge al di sotto dell'America meridionale

di David E. James

Lungo le coste occidentali dell'America meridionale, dal Venezuela e dall'America centrale fino all'estremità meridionale del Cile, si estende un'imponente serie di catene montuose, nell'insieme chiamata Cordigliera delle Ande. Chiaramente collegata alla Cordigliera, vi è una fossa oceanica lunga e arcuata, la fossa Perù-Cile, che corre approssimativamente parallela alle Ande da circa 4° di latitudine nord fino a circa 40° di latitudine sud. Tra il fondo della fossa e il piú alto picco delle Ande vi è un dislivello di circa 15 000 metri. L'arco andino, intendendo con tale termine sia il sistema montuoso sia la fossa connessa, è un sistema attivo e in evoluzione. Esso fa parte della cosiddetta cintura di fuoco circumpacifica e i vulcani attivi che punteggiano la cordigliera in tutta la sua lunghezza insieme con i disastrosi terremoti che puntualmente interrompono la vita di tutti i giorni dei sudamericani, sono un costante richiamo al fatto che ancora oggi sono molto attivi gli stessi processi orogenetici che hanno edificato la Cordigliera delle Ande.

Molti geologi, ed io fra quelli, ritengono che l'arco andino possa rappresentare un modello attuale di molte catene montuose piú antiche: se fossimo cioè vissuti nelle regioni della Sierra Nevada 100 milioni di anni fa o in quella degli Appalachi settentrionali 450 milioni di anni fa, avremmo assistito a un'attività orogenetica analoga. Fino a poco tempo fa, però, questo modo di vedere le cose non era condiviso da tutti; si riteneva infatti che i processi che avevano dato luogo a catene montuose del passato non fossero attivi oggigiorno. Questa supposizione sta oggi lasciando rapidamente il posto al concetto che l'attività orogenetica si sia sviluppata in relazione a fenomeni geologici collegati che si sono ripetuti molte volte nel corso almeno degli ultimi due miliardi di anni della storia della Terra e che questi continuino anche oggi. Considerando le cose da questo punto di vista, le Ande sono il piú importante esempio attuale di quei processi orogenetici attivi; cosí, attraverso lo studio delle Ande, si può cercare di capire come le antiche catene montuose siano nate, divenute mature e, in qualche caso, addirittura scomparse molto prima che l'uomo arrivasse a registrarne l'origine o lo sviluppo.

I concetti su cui si basava l'orogenesi sono stati di recente rivoluzionati dalla teoria della tettonica a zolle (si veda l'articolo *La tettonica a zolle crostali* di J.F. Dewey, in «Le Scienze», n. 48, agosto 1972). Questa teoria, che si è sviluppata con una rapidità impressionante nel corso degli ultimi 10 anni, si basa sul fatto che l'involucro piú esterno della Terra, la litosfera, è costituito da un mosaico di zolle rigide in movimento relativo l'una rispetto all'altra. Queste zolle, dello spessore di circa un centinaio di chilometri, comprendono non soltanto la crosta solida della Terra, ma anche parte del mantello superiore piú denso. I margini di zolla raramente coincidono con i margini continentali, cosicché i movimenti relativi di parti della superficie terrestre vengono ora considerati nei termini piú generali dei movimenti di zolla che non della cosiddetta deriva dei continenti. I continenti, che una volta si pensava navigassero attraverso la crosta oceanica come superbe navi, sono ora ridotti allo stato di passeggeri passivi posati su zolle litosferiche. Lungo le dorsali oceaniche il magma che fluisce dalle profondità del mantello terrestre, genera le zolle litosferiche. La litosfera di neoformazione si sposta allontanandosi dalle dorsali per lasciare il posto a sempre nuove quantità di magma che vengono iniettate lungo gli assi della dorsale. Tali zolle in espansione vengono consumate in corrispondenza delle fosse oceaniche, poiché là esse piegano verso il basso e si immergono nuovamente nel mantello da cui sono nate.

La maggior parte dell'attività tettonica mondiale – terremoti, vulcani e orogenesi – è concentrata lungo i margini di zolla. La costa occidentale dell'America meridionale coincide con uno di tali margini (*si veda l'illustrazione a pag. 44*): la zolla oceanica di Nazca, generata lungo la Dorsale pacifica orientale, viene distrutta nella fossa Perù-Cile, in corrispondenza della quale essa piega verso il basso e scivola al di sotto della zolla sudamericana a una velocità di circa 6 cm all'anno. L'analisi della struttura e della geologia delle Ande centrali rivela che la maggior parte dell'attività orogenetica attuale e dell'evoluzione geologica verificatasi negli ultimi 200 milioni di anni, può essere spiegata attraverso il fenomeno della subduzione della zolla oceanica al di sotto dell'America meridionale. Questo tipo di interazione fra le due zolle spiega anche l'accartocciamento del margine continenta-

Sulla pagina a fronte, la regione delle Ande presso il lago Titicaca, al confine tra Perú e Bolivia, in una fotografia scattata dall'ERTS (Earth Resources Technology Satellite). Il lago è sull'altipiano a circa 4 chilometri di altitudine sul livello del mare. L'altipiano è dominato a ovest (*in basso*) dalle montagne vulcaniche della cordigliera occidentale, a est (*in alto*) dalle montagne a pieghe della cordigliera orientale.

le stabile a formare catene di montagne a pieghe che ora costituiscono i contrafforti orientali delle Ande; spiega la nascita della grande cordigliera vulcanica andina a occidente e l'accrescimento continentale nella parte occidentale dell'America meridionale.

Oltre una decina di anni fa, la Carnegie Institution di Washington, con la collaborazione di un certo numero di istituzioni sudamericane, formulò un programma di ricerca che aveva lo scopo di studiare la natura e l'evoluzione delle Ande centrali del Perù meridionale, della Bolivia e del Cile settentrionale. Parecchi anni di lavoro hanno portato a una conoscenza abbastanza completa delle proprietà fisiche della Terra al di sotto delle Ande centrali; solo di recente, però, sono stati forgiati gli strumenti concettuali attraverso i quali comprendere le forze e il succedersi degli eventi orogenetici che hanno determinato la formazione delle Ande e che sono tutt'oggi all'origine del vulcanismo e della sismicità che si registrano lungo il sistema montuoso.

Per comprendere l'evoluzione delle Ande centrali è necessario soffermarsi su due tipi di diversi dati di fatto, quelli geofisici e quelli geologici. Fra quelli geofisici vi sono la distribuzione dei terremoti e la distribuzione entro la crosta e il mantello superiore di alcune proprietà fisiche: la velocità delle onde sismiche, l'assorbimento della loro energia (cioè la loro attenuazione) e la densità delle rocce. Da queste proprietà si possono trarre informazioni sul tipo di rocce che costituisce l'interno della Terra e sulla loro maggiore o minore rigidità. I dati geologici derivano dallo studio dei tipi di rocce e delle loro strutture cosí come si possono rilevare alla superficie terrestre; soltanto sulla base di questi dati si può ricostruire la storia di eventi accaduti nel passato. Dapprima prenderemo in considerazione i dati geofisici, poiché essi forniscono la base d'osservazione per la maggior parte di ciò che sappiamo sul margine lungo il quale vengono a contatto la zolla sudamericana e la zolla di Nazca.

La chiave per ricostruire le interazioni fra zolle diverse è fornita dalla distribuzione dei terremoti. È principio generale della teoria della tettonica a zolle che i piani inclinati lungo i quali si distribuiscono i fuochi dei terremoti al di sotto delle fosse e degli archi insulari, chiamati piani di Benioff (dal nome del sismologo Hugo Benioff), delineino la superficie superiore di una zolla oceanica in via di sprofondamento. Le zolle litosferiche in via di sprofondamento sono piú fredde e quindi piú rigide dell'astenosfera (parte del mantello caldo) nella quale affondano. È opinione comune che soltanto nella litosfera vi siano rocce abbastanza rigide da generare, spezzandosi, dei terremoti: questa supposizione è avallata dal fatto che nella maggior parte degli archi insulari, i terremoti sembrano avere origine quasi esclusivamente all'interno della zolla in via di sprofondamento.

Al di sotto delle Ande centrali, tuttavia, i terremoti avvengono non soltanto nella zolla discendente, ma anche in un cuneo continuo tra la superficie terrestre e il tetto della zolla oceanica in via di sprofondamento, cuneo che si estende fino a una profondità compresa tra i 200 e i 300 chilometri (*si veda la figura in alto a pag. 50*). Questa osservazione potrebbe ingenerare il sospetto che la zolla litosferica sudamericana, che sovrascorre su quella pacifica, possa avere uno spessore anormale di 200 o 300 chilometri, che sarebbe molto maggiore dei normali 100 chilometri di spessore osservati per la maggior parte delle altre zolle e comunque in stridente contrasto con lo spessore di 50 chilometri misurato con metodi sismici per la zolla di Nazca che le sottoscorre. Una indicazione piú diretta del fatto che la litosfera al di sotto delle Ande ha uno spessore compreso tra 200 e 300 chilometri, proviene dai risultati della sismica che mostrano che in quella regione non vi è una zona a bassa velocità o una zona di elevata attenuazione delle onde sismiche.

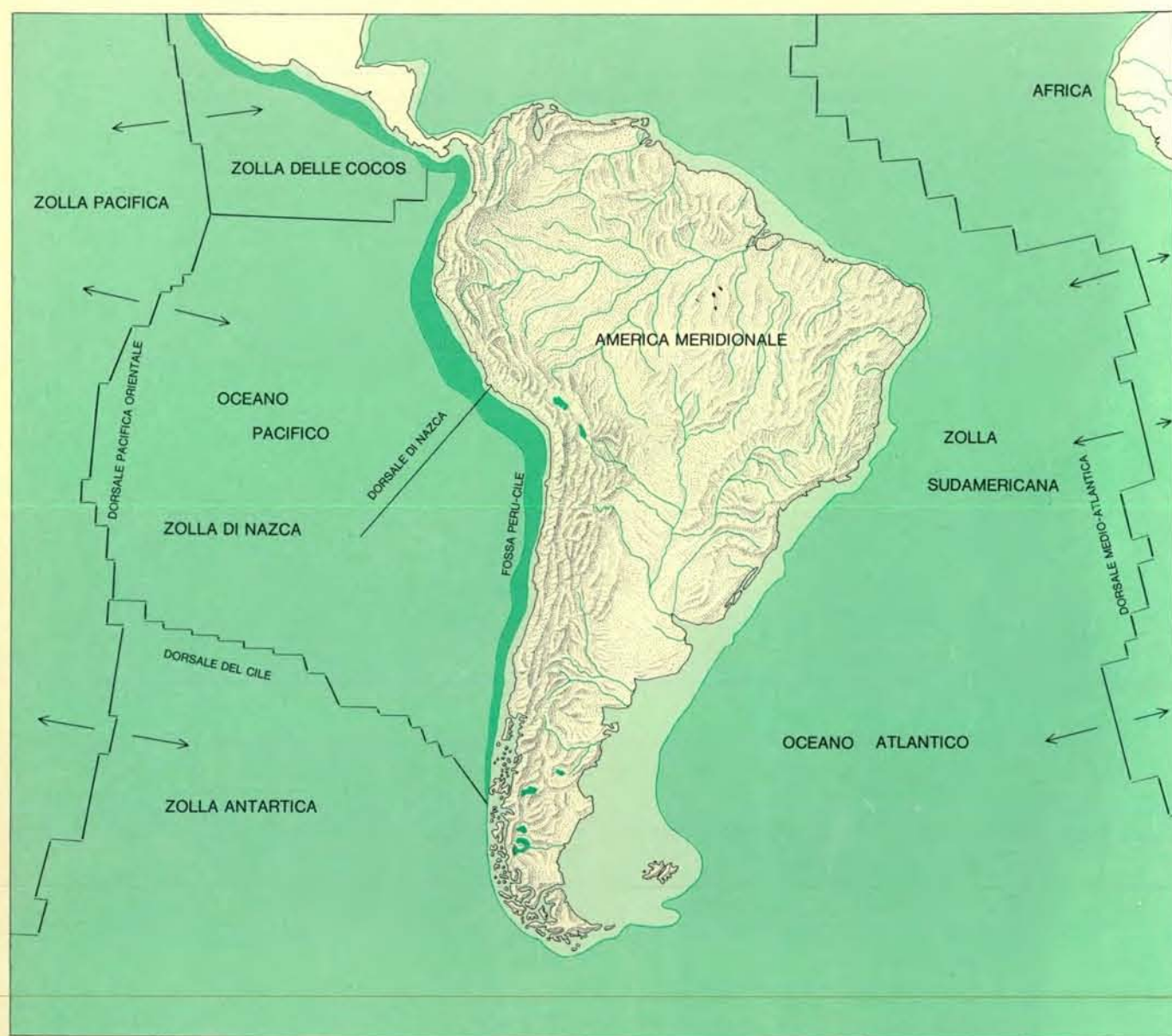
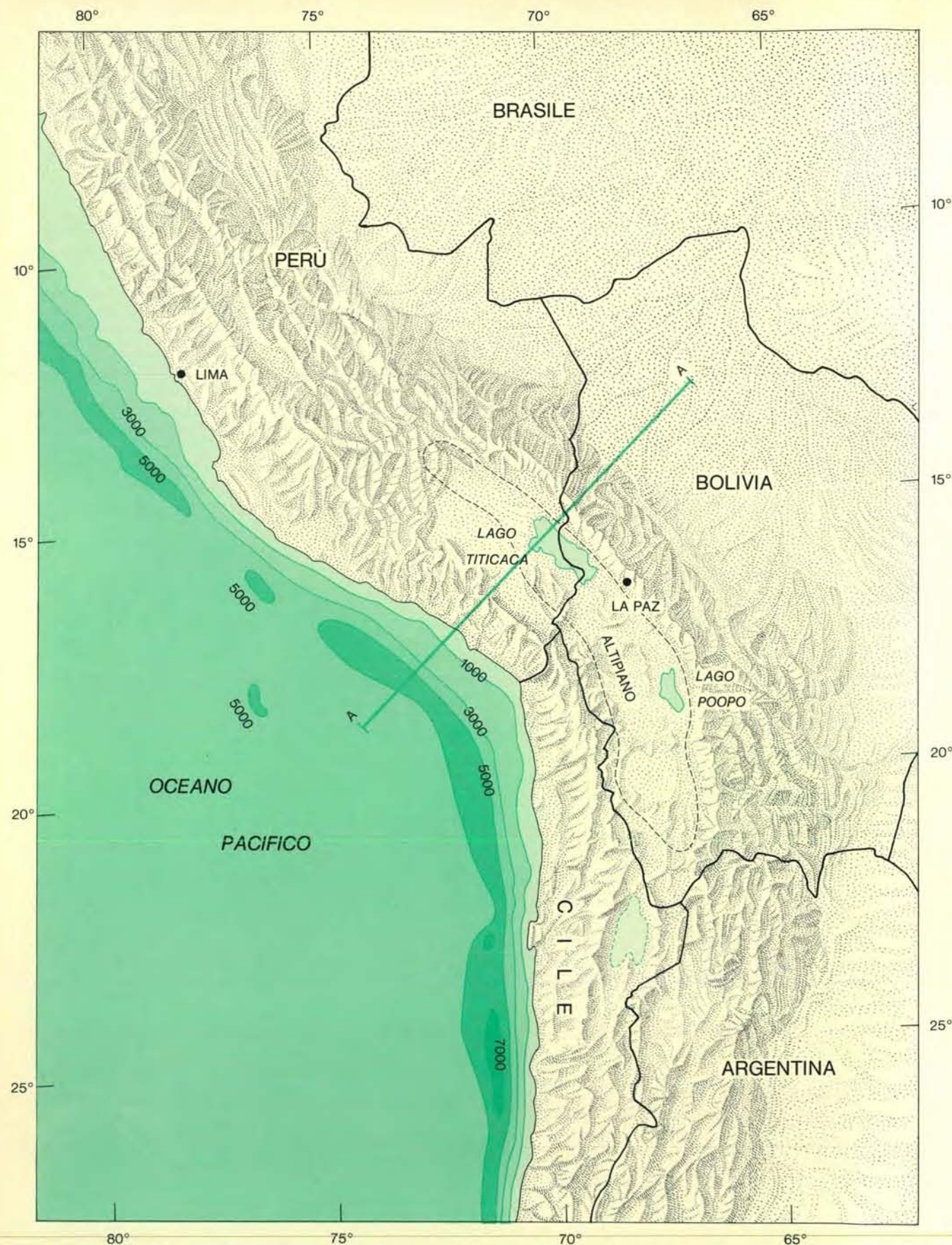
Sia la zona a bassa velocità sia quella a elevata attenuazione sono normalmente interpretate come regioni nelle quali appaiono rocce meno rigide e sono di solito associate all'astenosfera. Per contro, la loro assenza implica la conclusione che le rocce sono comparativamente rigide. Una conseguenza importante dell'elevato spessore della litosfera al di sotto delle Ande è che questo impedisce l'inversione dei moti convettivi nel mantello al di sopra del piano di Benioff. A questa inversione è attribuita la genesi dei centri secondari di espansione dei fondali oceanici che si sviluppano dietro agli archi insulari e che invece mancano del tutto dietro l'arco continentale andino.

Un secondo e importante gruppo di osservazioni geofisiche, relativo allo spessore della crosta, fornisce una ulteriore indicazione del legame esistente tra il comportamento attuale delle zolle e la storia orogenetica delle Ande cosí come può essere ricostruita sulla base dei dati geologici. I dati sismici mostrano che la crosta delle Ande centrali ha uno spessore molto variabile: su una distanza di poco piú di 500 chilometri essa cambia da circa 11 chilometri (compresa l'acqua) nel bacino del Pacifico a circa 30-35 chilometri lungo la costa (uno spessore considerato normale per la crosta continentale), a piú di 70 chilometri al di sotto della dorsale vulcanica della catena occidentale, a circa 50-55 chilometri sotto le catene orientali a pieghe e, da ultimo, a circa 35 chilometri al di sotto dello scudo brasiliano. Fuori delle Ande, solo nell'Himalaya sono stati osservati spessori della crosta fino a 70 chilometri.

Questi risultati, ottenuti attraverso misure sismiche dirette, potevano essere previsti semplicemente sulla base della topografia e del principio dell'isostasia. Questo principio stabilisce che un eccesso di massa al di sopra del livello del mare – per esempio una catena montuosa – deve essere compensato da un equivalente difetto di massa in profondità: le radici crostali delle catene montuose (che coinvolgono rocce del mantello) devono essere meno dense delle rocce che le catene stesse coinvolgono. Il principio dell'isostasia è, naturalmente, valido anche se formulato all'inverso: ogni difetto di massa in profondità deve essere compensato da un eccesso di massa in superficie; a ogni aggiunta o ispessimento della crosta, come quelli che si verificano per iniezione di magma dal basso, deve corrispondere un sollevamento in superficie. È noto infatti, da studi sismici, che nelle zone che per lungo tempo si sono mantenute presso il livello del mare (e che quindi si suppone siano in equilibrio isostatico) la crosta continentale ha spessori di 30-40 chilometri. Le aree al di sotto del livello del mare hanno una crosta piú sottile, quelle al di sopra hanno una crosta piú spessa. Questo fatto diviene della massima importanza quando si consideri l'evoluzione crostale, dato che la geologia indica che la maggior parte delle regioni piú elevate delle Ande si trovavano un tempo presso il livello del mare.

Un'altra proprietà geofisica, la velocità sismica, può essere utilizzata per stabilire la composizione delle rocce entro la crosta andina e in quella oceanica adiacente. La velocità sismica di molte rocce comuni, in diverse condizioni di temperatura e pressione, è stata misurata attraverso esperimenti di laboratorio. Le velocità sismiche di rocce nella crosta oceanica inferiore sono caratteristiche di una roccia di composizione basaltica sottoposta a quei valori di temperatura e pressione. Un fatto interessante è che, anche se pressione e temperatura nelle radici crostali delle Ande sono molto maggiori che non quelle nella crosta oceanica inferiore, le velocità sismiche delle rocce sono le stesse. Nello stesso tempo si sa che, se una roccia basaltica fosse sottoposta alle pressioni e temperature vigenti in profondità nella crosta continentale, essa dovrebbe subire una trasformazione metamorfica che la muterebbe in un tipo di roccia caratterizzata da densità e velocità sismiche piú elevate di quanto non si osservi.

Quando si esamini in dettaglio i diversi tipi di roccia, soltanto le rocce crostali piú leggere, simili cioè come composizione a quelle che si osservano in superficie, mostrano di raggiungere, alla temperatura e pressione vigenti in profondità nella crosta, le velocità sismiche che in realtà si osservano. Ciò conduce alla conclusione che la crosta andina sia abbastanza omogenea verticalmente e che le rocce nelle radici delle Ande siano simili, per composizione, alle rocce vulcaniche e plutoniche che costituiscono i livelli superiori della crosta. Eccezioni all'omogeneità verticale della crosta si verificano nell'altipiano e nelle cordigliere orientali. Le variazioni nella velocità sismica indicano che l'altipiano potrebbe trovarsi alla sommità di una pila di circa 30 chilometri di rocce sedimentarie forse mescolate con depositi vulcanici e che le cordigliere orientali potrebbero svilupparsi su un accumulo dello spessore di almeno 5-10 chilometri di rocce sedimentarie.

Quando si vuole ricostruire la storia dello sviluppo del sistema andino è necessario ricorrere ai dati geologici. Fra le rocce piú antiche delle Ande centrali vi sono le rocce sedimentarie paleozoiche depositatesi tra 250 e 450 milioni di anni fa. Il tempo non è stato gentile con esse: le troviamo dislocate e accartocciate a formare le catene a pieghe dei contrafforti orientali. Queste rocce, circa 10 chilometri di una monotona alternanza di strati

Situazione geotettonica delle Ande. Il margine tra la zolla di Nazca e quella sudamericana costituisce il piú importante esempio di interazione tra una zolla oceanica e una continentale. La zolla di Nazca si genera lungo la Dorsale pacifica orientale e si consuma nella fossa Perú-Cile. L'interazione delle due zolle ha determinato la deformazione del margine dell'America meridionale da cui è derivata la formazione del sistema andino. Le frecce indicano l'espansione rispetto alle dorsali.

Le Ande centrali, una regione in cui sembra che le forze orogenetiche siano ancora attive. A settentrione di Lima, le Ande formano un sistema unico costituito da catene montuose strettamente associate che corrono parallele alla costa. A meridione di Lima, invece, il sistema si ramifica e la catena orientale a pieghe corre per centinaia di chilometri verso l'interno mentre la catena vulcanica occidentale continua a svilupparsi parallelamente alla costa. Tra i due rami si trova l'altipiano, un'estesa pianura che si sviluppa alla sommità di un enorme prisma di detriti sedimentari erosi dalle catene adiacenti. Nel Cile settentrionale i due rami si ricongiungono nuovamente. La linea A-A' indica la regione schematizzata in alcune delle figure che seguono.

sabbioso-argillosi, sono chiamate depositi geosinclinalici (si veda l'articolo *Geosinclinali, orogenesi e accrescimento dei continenti*, di Robert S. Dietz, in «Le Scienze», n. 46, giugno 1972). Un tempo esse formavano il tranquillo margine continentale paleozoico del Sudamerica occidentale. Un margine attuale analogo è il margine continentale inattivo della costa orientale del Nordamerica, dove, tra la piattaforma continentale e il piano abissale dell'oceano, si è formato un enorme cuneo sedimentario largo oltre 250 chilometri e spesso oltre 10.

Nel Permiano e nel Triassico, tra circa 200 e 250 milioni di anni fa, il margine occidentale del Sudamerica, prima tranquillo, fu gradualmente sconvolto dalle prime fasi della fratturazione del supercontinente di Pangea e dall'instaurarsi del ciclo tettonico ora in atto. Il cuneo continentale divenne instabile e i depositi geosinclinalici vennero deformati e sollevati. Nell'accumulo sedimentario si intrusero magmi forse derivati dalla fusione parziale dei depositi geosinclinalici più profondi a formare batoliti, cioè imponenti corpi di rocce intrusive in profondità e vulcani in superficie. I vulcani antichi sono da lungo tempo scomparsi ma, intercalati alle rocce sedimentarie, si trovano depositi vulcanici di questo periodo. Le rocce intrusive che probabilmente rappresentano i bacini magmatici che alimentavano i vulcani, si trovano comunemente affioranti oggi al nucleo delle catene orientali a pieghe, avvolte da rocce sedimentarie spiegazzate.

Circa 190 milioni di anni fa, cioè agli inizi del Giurassico, l'asse principale dell'attività tettonica si spostò di alcune centinaia di chilometri verso ovest, cioè in direzione dell'oceano. La litosfera si spezzò a quei tempi lungo il margine che separava il continente sudamericano dal bacino del Pacifico e la zolla oceanica incominciò il suo sprofondamento nel mantello al di sotto del margine occidentale del Sudamerica. Quanto alle ragioni che provocarono la rottura catastrofica della litosfera, è possibile formulare solo delle ipotesi, anche se vi sono indicazioni del fatto che ciò può essere avvenuto in relazione con l'inizio dell'espansione dei fondali oceanici lungo un'ancestrale Dorsale pacifica orientale. Sappiamo comunque per certo che l'oceano Atlantico meridionale non aveva ancora cominciato ad aprirsi e che Sudamerica e Africa erano ancora unite.

Come la zolla pacifica si immerse al di sotto del Sudamerica, giunsero in superficie magmi andesitici trasudati dalla crosta oceanica basaltica in via di sprofondamento. L'andesite è una roccia vulcanica caratteristica delle Ande: è più ricca in silice, e quindi meno densa, del basalto; la sua composizione è intermedia tra quella di un basalto e quella di un granito. Le rocce andesitiche giurassiche sono ben conservate lungo le regioni costiere del Perù meridionale e del Cile settentrionale. È tuttavia ancora non del tutto chiaro se le andesiti derivino da una crosta sialica (cioè continentale) o simatica (cioè oceanica). Le lave stesse sembrano essere fluite in ambiente subacqueo dato che molte di esse mostrano i segni di una profonda alterazione operata da acqua salata.

Da tutto questo si potrebbe dedurre che l'arco vulcanico delle Ande centrali cominciò come un arco insulare nell'oceano al largo delle coste degli antichi Perù meridionale e Cile settentrionale (*si veda la figura alle pagine 48 e 49*). Ciononostante il quadro dell'arco insulare giurassico non è chiaro poiché le rocce vulcaniche giurassiche nel Perù meridionale sono incuneate entro rocce cristalline metamorfiche di almeno 400 milioni di anni di età. Cosa rappresentino questi resti di una antica crosta sialica a 300 chilometri a ovest dei luoghi in cui affiorano normalmente i depositi geosinclinalici paleozoici è ancora un mistero. La presenza, tuttavia, di rocce sialiche non implica necessariamente che l'arco giurassico si sia formato su crosta continentale. Queste rocce potrebbero avere fatto parte, nel Paleozoico, di un microcontinente o di una penisola che si estendeva dalla costa sudamericana verso occidente; oppure potrebbero essere i resti di una zattera di crosta sialica, galleggiante al di sopra della zolla oceanica, portata contro la zolla sudamericana e appiccicatasi a quest'ultima mentre la prima sprofondava verso l'interno della Terra in corrispondenza della fossa.

A parte queste difficoltà, è chiaro che la formazione dell'arco vulcanico andino era assai avanzata nel Giurassico. Il fatto stesso che le rocce vulcaniche giurassiche siano tuttora largamente conservate indica che l'arco vulcanico giurassico non è mai venuto a trovarsi a un'altitudine eccessiva sul livello del mare e che quindi è sfuggito a un'azione demolitrice profonda da parte dell'erosione. Sembra inoltre che l'intera regione a oriente dell'arco giurassico, fino a dove almeno si estendono le Ande attuali, deve essere stata prossima al livello del mare, dato che in tutta la zona sono segnalati depositi sedimentari marini. Questa osservazione è importante perché mostra che la crosta delle Ande era ancora sottile, probabilmente non più spessa di 35 chilometri.

Si è notato che l'arco giurassico si è formato in un periodo in cui Sudamerica e Africa costituivano ancora un continente unico. La datazione delle anomalie magnetiche dei fondali dell'Atlantico meridionale, effettuata da Walter C. Pitman III e altri del Lamont-Doherty Geological Observatory della Columbia University, indica che l'Atlantico meridionale ha cominciato ad aprirsi soltanto attorno a 135 milioni di anni fa per l'attività intervenuta lungo quella porzione della Dorsale medio-atlantica. La formazione dell'arco giurassico inaugurò l'imponente serie di episodi orogenetici che hanno prodotto il sistema delle Ande, ma l'orogenesi andina cominciò seriamente soltanto qualche tempo dopo la separazione dell'America meridionale dall'Africa.

Circa 100 milioni di anni fa, nel Cretaceo, parallelo all'arco giurassico e tra questo e il continente, cominciò a formarsi un secondo arco vulcanico consistente (questa trattazione dell'evoluzione delle Ande sotto forma di una serie di pochi episodi distinti è chiaramente una semplificazione ma è utile per tracciare la storia del sistema montuoso; in realtà, l'attività eruttiva è stata abbastanza continua da circa 200 milioni di anni fa fino ai giorni nostri). Le lave di questa catena vulcanica fluirono al di sopra del livello del mare, una chiara prova che i vulcani si trovavano al di sopra della crosta continentale. Tra 50 e 60 milioni di anni fa, con l'invasione della crosta da parte di imponenti quantità di magma, l'attività lungo l'arco cretaceo raggiunse il suo massimo. I magmi cristallizzarono a formare gli enormi batoliti che ora affiorano lungo tutti i contrafforti occidentali delle catene occidentali.

I vulcani di allora, che dovevano innalzarsi parecchio al di sopra del livello del mare, sono stati da lungo tempo distrutti dall'erosione, scoprendo le radici intrusive dell'arco vulcanico: sono i batoliti andini. In alcune zone, come nel Perù meridionale, la erosione non ha agito così in profondità, cosicché piccoli batoliti sono ancora coperti da depositi vulcanici di età e composizione simili. Questi piccoli corpi intrusivi fanno probabilmente parte dei camini vulcanici che alimentavano gli apparati superficiali. Col tempo, anche i batoliti verranno asportati, lasciando affiorare soltanto le radici intrusive più profonde dell'arco vulcanico cretaceo del quale non rimarrà altra traccia.

Il rigonfiamento crostale che accompagnò l'intrusione del magma produsse danni importanti nelle rocce geosinclinaliche facilmente deformabili delle catene orientali. I magmi ascendenti provocarono una dilatazione della crosta in corrispondenza dell'arco cretaceo. Le rocce geosinclinaliche furono alternativamente rimosse e spinte di lato dalla catena nascente e furono spiegazzate e incluse nelle catene orientali a pieghe. Dalle adiacenti catene montuose furono erose enormi quantità di sedimenti che poi venivano depositati nel bacino corrispondente all'attuale altipiano che, in alcune zone, ricevette, nel solo Cretaceo, piú di 7 chilometri di sedimenti.

È significativo il fatto che l'arco cretaceo si sia sviluppato sul lato continentale del piú antico arco giurassico, poiché si ritiene che, normalmente, all'atto in cui i frammenti crostali vengono trasportati dalla crosta oceanica fino al di sotto degli archi vulcanici, fossa e arco migrino verso l'oceano. Nella zona andina è invece vero il contrario: nel corso dell'evoluzione delle Ande, l'asse dell'arco vulcanico si è sempre spostato verso l'entroterra allontanandosi dal mare. Anche alla scala del dettaglio si trovano intrusioni successive dei batoliti andini che si spostano in direzione del continente. Si possono formulare diverse ipotesi sulle ragioni di questa migrazione verso est, ma torneremo su questo problema quando prenderemo in considerazione le origini delle imponenti quantità di magma dalle quali si è formata la crosta delle Ande.

Gli edifici vulcanici andini attuali cominciarono a formarsi circa 15 milioni di anni fa. Nel Cile settentrionale e nel Perù meridionale, profonde fessure e diversi centri eruttivi vomitarono con violenza esplosiva imponenti masse di ceneri vulcaniche silicee, che finirono col coprire centinaia di migliaia di chilometri quadrati. Anche ora, quando la maggior parte di questi depositi piroclastici è stata da tempo asportata dall'erosione, si stima che i loro resti coprano da 10 000 a 15 000 chilometri quadrati per uno spessore medio di 500 metri. Le eruzioni di ceneri continuarono a un ritmo pressoché stazionario fino a circa 4 milioni di anni fa, quando esse cessarono improvvisamente e furono subito seguite dall'eruzione di lave andesitiche. Queste ultime formano i grandi vulcani a strato, alcuni dei quali ancora attivi, che dominano la Cordigliera, e che nelle Ande centrali superano largamente i 6 chilometri di altezza.

Il massiccio afflusso di magma nella crosta sottostante all'attuale dorsale vulcanica andina, provocò un rigonfiamento della crosta e un'estesa serie di dislocazioni nella zona dell'altipiano e nelle catene orientali. I sedimenti geosinclinalici furono ulteriormente spinti di lato dall'espandersi della cintura magmatica e le catene orientali furono schiacciate a formare catene montuose alte e strette. Detriti sedimentari, provenienti dalle catene orientali e occidentali, andarono accumulandosi nel Cenozoico nella zona dell'altipiano fino a raggiungere spessori superiori ai 15 chilometri. Ora stiamo assistendo finalmente a una temporanea diminuzione di attività e dal rilassamento delle forze di compressione che ne deriva si formano strutture di distensione nell'altipiano. Il lago Titicaca è sede di una struttura di questo tipo: è infatti un graben, cioè una lunga depressione compresa fra faglie di distensione.

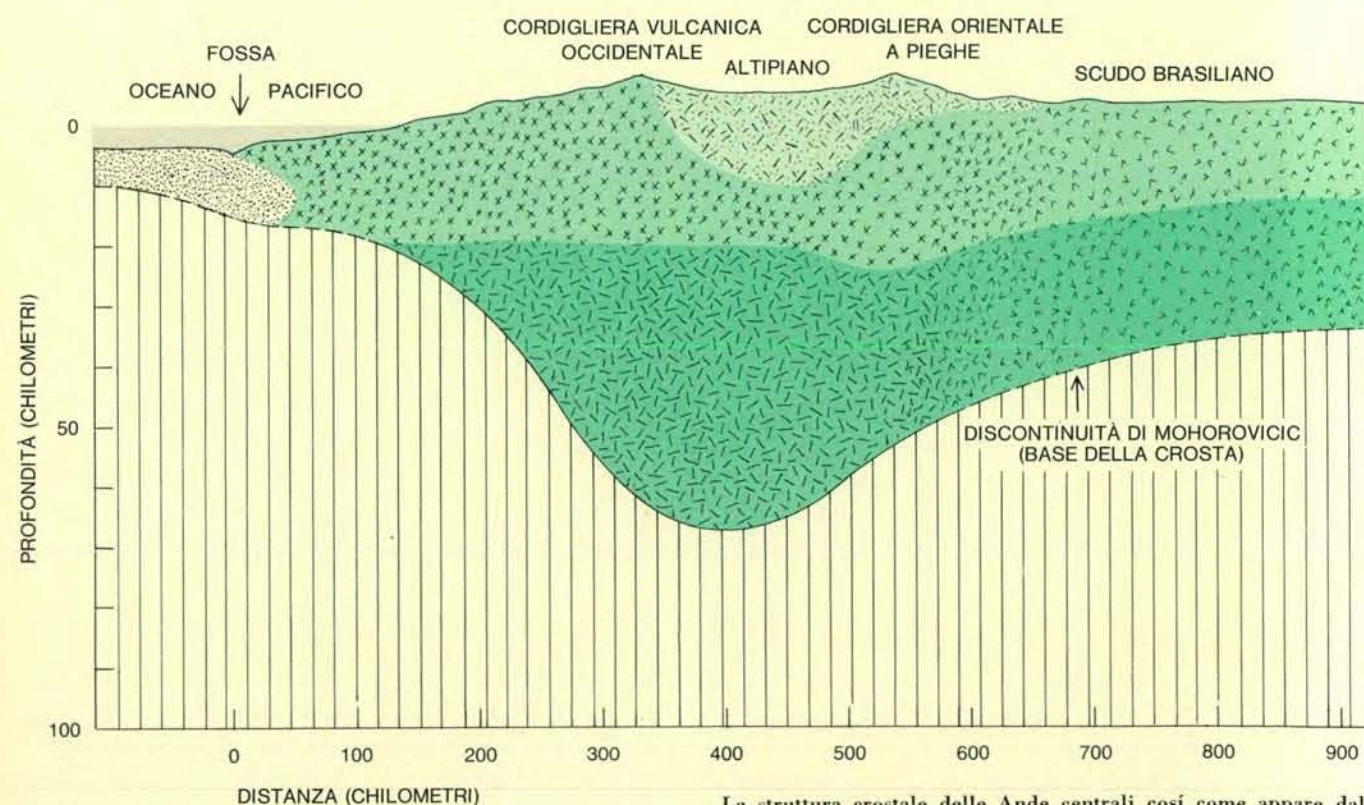
Fin qui abbiamo considerato l'evoluzione delle Ande per quanto se ne può sapere dalla geologia di superficie, descrivendo il progressivo spostamento dell'arco vulcanico verso l'entroterra, il conseguente ripiegamento dei sedimenti geosinclinalici paleozoici e il graduale sollevamento della catena delle Ande. Non si è invece ancora parlato dello sviluppo della crosta e dell'origine delle rocce crostali che si trovano al di sotto della superficie.

I concetti orogenetici classici, formulati ben prima di quelli della tettonica a zolle, stabilivano che le rocce geosinclinaliche si deponevano in bacini allungati sede di intensa subsidenza; tali bacini erano il luogo dal quale si sviluppavano poi le catene montuose che si supponeva nascessero dalla deformazione e fusione degli imponenti depositi sedimentari geosinclinalici. Rimaneva comunque un mistero il meccanismo attraverso il quale tali rocce sedimentarie venivano fuse e deformate, cosí come era un problema l'origine del bacino subsidente stesso. Quando si applichino i concetti classici delle geosinclinali alle Ande centrali, emerge un certo numero di problemi. La crosta al di sotto della dorsale vulcanica ha uno spessore di oltre 70 chilometri e ciononostante nessun bacino sedimentario può essersi mantenuto per lungo tempo al di sopra di una crosta di spessore superiore ai 35-40 chilometri senza trovarsi a emergere e quindi senza trovarsi di fronte all'interruzione del processo di sedimentazione.

Se la crosta della cordigliera occidentale è costituita da rocce geosinclinaliche rimobilizzate, queste ultime devono aver subito un raccorciamento tale da produrre il raddoppio dello spessore della crosta. Ma nella dorsale vulcanica si trovano poche strutture come faglie di compressione e pieghe rovesciate, tipici effetti del raccorciamento crostale. Al contrario, studi sismici e tettonici indicano che lo stile di deformazione è di tipo distensivo; la crosta al di sotto dell'arco vulcanico non si sta contraendo ma piuttosto distendendo. Chiaramente, ciò che si sa della crosta delle Ande non può facilmente essere giustificato con i concetti classici dell'orogenesi.

Solamente l'avvento della teoria della tettonica a zolle ha sistemato questa infelice discrepanza tra le monta-

Evoluzione delle Ande a nord-ovest della linea A-A' indicata nella figura a pagina 46. Nel Paleozoico superiore (*schema 1*), circa 250 milioni di anni fa, l'attuale linea di costa era ricoperta da depositi sedimentari e non vi erano montagne. Nel Triassico e nel Giurassico (*schema 2*), circa 200 milioni di anni fa, la litosfera iniziò a sottoscorrere deformando le rocce sedimentarie e spingendole verso l'alto e verso est. Il magma derivato dalla zolla oceanica in via di sprofondamento, piú leggero, risalí generando un arco vulcanico nelle acque costiere del Sud America occidentale. Negli strati sedimentari delle catene orientali si formarono alcuni batoliti o ammassi di rocce intrusive. Nel Cretaceo e all'inizio del Cenozoico (*schema 3*), cioè tra 100 e 60 milioni di anni fa, a occidente dell'arco giurassico, cominciò a formarsi un secondo arco vulcanico. Il rigonfiamento nella crosta, provocato dall'ascesa del magma, ha spinto di lato le rocce sedimentarie antiche che si sono accavallate a formare le catene a pieghe della cordigliera orientale. Il materiale eroso da tali montagne si è accumulato a formare la regione dell'altipiano. La formazione dell'attuale catena è iniziata tra 10 e 15 milioni di anni fa, raggiungendo nel Pliocene o nel Pleistocene (*schema 4*), cioè uno o due milioni di anni fa, l'attuale struttura.

Le differenze strutturali tra l'arco vulcanico delle Ande e l'arco vulcanico delle Tonga-Figi giustificano la diversità sia delle loro forme sia della distribuzione dei terremoti. Nell'arco delle Tonga la maggior parte dei terremoti si sviluppano all'interno della zolla discendente che è inclusa nel materiale astenosferico; a ovest dell'arco vi sono, inoltre, indicazioni di fenomeni di minore entità di espansione dei fondali oceanici in atto. Tali osservazioni sembrano indicare che l'attrito derivato dalla zolla discendente determini un rovesciamento nei moti convettivi nell'astenosfera sovrastante, che, a sua volta, causa l'espansione dietro l'arco. I terremoti nelle Ande non sono limitati alla zolla discendente, né vi è alcuna indicazione di espansione crostale a oriente dell'arco vulcanico. Con ogni probabilità, la spessa litosfera al di sotto del margine occidentale del Sud America impedisce la convezione e fenomeni di espansione crostale secondari, cosicché le forze generate dalla zolla discendente producono terremoti invece che moti convettivi. L'intensità dell'attività sismica sotto le Ande è indicata dalla densità dei punti.

La struttura crostale delle Ande centrali così come appare dall'analisi della propagazione delle onde sismiche attraverso le diverse rocce della crosta. Il colore chiaro indica una velocità sismica di circa 5 chilometri al secondo; il colore di media intensità, una velocità di circa 6 chilometri al secondo; il colore intenso, circa 6,6 chilometri al secondo. Le parti più spesse della crosta si trovano al di sotto della cordigliera occidentale e della parte occidentale dell'altipiano, ove sono penetrate nella crosta grandi quantità di magma proveniente da regioni più interne della Terra. La crosta della cordigliera orientale a pieghe è stata dislocata e compressa fino a raggiungere uno spessore di 50 chilometri. Al di sotto dell'altipiano giacciono circa 30 chilometri di sedimenti erosi dalle montagne circostanti.

gne e le teorie orogenetiche: essa ha proposto meccanismi orogenetici e processi che sono in relazione evidente con la fascia orogenica andina. Appare ora evidente (anche se infuria ancora qualche scaramuccia scientifica di retroguardia) che le rocce sedimentarie geosinclinaliche sono soltanto vittime dei processi tettonici che accompagnano la subduzione, intrappolate lungo i margini continentali. Il cuneo geosinclinalico deve parteciparvi ma non può mai essere il motore del processo orogenetico stesso. Se questa analisi è esatta ne segue che i magmi che edificano la dorsale vulcanica delle Ande centrali provengono dal di sotto della crosta e che le imponenti masse rocciose che sono state prodotte dalla solidificazione di questi magmi, hanno determinato un ispessimento crostale e un deciso sollevamento dell'arco.

Rimane allora il problema delle sorgenti del magma al di sotto della crosta. Vi sono fondamentalmente due possibilità: fusione parziale delle rocce dense del mantello nel cuneo che si trova tra la crosta e la superficie superiore della zolla in via di sprofondamento; oppure fusione parziale dei basalti e dei sedimenti della crosta oceanica che sprofondano nel mantello trascinati dal nastro trasportatore litosferico. La maggior parte delle osservazioni si oppone al fatto che le rocce andesitiche delle Ande possano avere una sorgente nel mantello. Se gli spessori della crosta nell'arco vulcanico si sono raddoppiati per semplice addizione di magma dal basso, per produrre il necessario incremento della crosta occorre far intervenire la fusione parziale di un 20% delle rocce del mantello tra la crosta e la zona di subduzione. Studi sperimentali sulla fusione parziale di materiali analoghi a quelli che si suppone costituiscano il mantello, non favoriscono l'ipotesi che una tale quantità di magma andesitico possa essere estratta da rocce di quel tipo. La derivazione delle rocce andine dal mantello sembra possibile soltanto se un flusso dal mantello alimentasse costantemente la crosta con roccia completa ancora della sua frazione a basso punto di fusione. Ma, come si è visto, il mantello superiore al disotto delle Ande è costituito da litosfera rigida fino ad almeno 200 o 300 chilometri di profondità e, dunque, probabilmente immobile.

Attualmente è preferita una proposta alternativa che io stesso ritengo fornisca una migliore spiegazione della genesi della crosta andina: i magmi sarebbero derivati dalla crosta oceanica in via di sprofondamento. Le rocce della crosta oceanica hanno temperature di fusione di diverse centinaia di gradi centigradi inferiori a quelle del mantello. Come la crosta oceanica fredda scende nel mantello, viene riscaldata dal mantello caldo che la avviluppa e, a profondità di circa 100-150 chilometri, la crosta basaltica può ragionevolmente raggiungere il suo punto di fusione. La frazione fusa della crosta oceanica è meno densa sia della roccia madre sia delle rocce del mantello che si trovano tutt'attorno; essa si trova quindi nelle condizioni di migrare verso l'alto nella crosta. La maggior parte del magma rimarrebbe intrappolato e solidificherebbe a livelli bassi e medi della crosta; soltanto una piccola quantità troverebbe la strada per giungere in superficie attraverso i vulcani.

Un fenomeno connesso con le origini delle rocce crostali è quello della migrazione verso est dell'attività eruttiva. Ponendo che le rocce abbiano origine attraverso la fusione parziale di una zolla in via di sprofondamento, vi sono diversi modi di spiegare tale fenomeno. Se, per esempio, la profondità alla quale avviene la fusione rimane costante, allora tanto una progressiva diminuzione nell'inclinazione della zolla discendente quanto una migrazione progressiva della zolla verso il continente, potrebbero causare una migrazione verso est dell'attività eruttiva. Un'altra possibilità: se la posizione della zolla discendente resta costante, la profondità alla quale avviene la fusione della zolla può aumentare con il tempo e ciò provocherebbe lo spostamento della sorgente del magma più in profondità e quindi verso est. Una considerazione fondamentale è, a questo punto, quella della posizione e inclinazione della zolla discendente dal momento in cui la subduzione ha avuto inizio. L'inclinazione del piano di Benioff è mutata negli ultimi 200 milioni di anni o si è spostata la fossa?

Una risposta parziale a questo interrogativo si può dare attraverso un esame del chimismo delle rocce stesse. William R. Dickinson, della Stanford University e Trevor Hatherton, del New Zealand Department of Scientific and Industrial Research, hanno dimostrato l'esistenza di una chiara correlazione tra la quantità di potassa (K_2O) nelle rocce vulcaniche attuali e la profondità del sottostante piano di Benioff. Per alcuni tipi particolari di rocce, come le andesiti raccolte da archi insulari in tutto il mondo, il contenuto in potassa è più o meno una funzione lineare della profondità del piano di Benioff sottostante. Rovesciando questa relazione, è possibile stimare la profondità di piani di Benioff del passato semplicemente misurando il contenuto di potassa in rocce antiche.

Si comincia solo ora ad applicare questa tecnica di ricerca alle Ande. I dati non sono ancora completi, ma mentre si passa dalle rocce dell'arco giurassico a quelle dell'arco cretaceo fino a quelle dell'arco attuale, si riscontra un progressivo incremento nel contenuto in potassa. Questo incremento progressivo, naturalmente, è anche correlato con l'aumento di profondità rispetto all'attuale zona di Benioff. Variazioni analoghe si trovano anche per il contenuto in stronzio, un altro parametro chimico che ha mostrato una correlazione positiva con la profondità del piano di Benioff.

È perciò possibile formulare l'ipotesi che la posizione del piano di Benioff non sia radicalmente mutata nel tempo. Se così è, la migrazione verso est dell'arco, avvenuta negli ultimi 200 milioni di anni, implica che la posizione del mantello nella quale avviene la fusione della crosta oceanica sia stata spinta progressivamente a sempre maggiori profondità. Ciò potrebbe trovare una giustificazione nel fatto che il continuo sottoscorrimento di crosta oceanica fredda nel mantello al di sotto del continente, abbia gradualmente raffreddato i materiali che costituiscono il mantello stesso: in questo modo la temperatura alla quale avviene la fusione nella zolla discendente verrebbe raggiunta a profondità via via crescenti. Se un processo di questo tipo giustifica la migrazione dell'attività eruttiva dalla fossa verso l'entroterra, allora, continuando la subduzione al di sotto delle Ande centrali, le dorsali vulcaniche del futuro coinvolgeranno sempre più le rocce geosinclinaliche antiche, spingendole sempre più a fondo nell'entroterra.

Un giorno, tuttavia, l'orogenesi andina si spegnerà. La catena vulcanica smetterà di crescere e avrà inizio l'inevitabile demolizione da parte dell'erosione. Per ridurre le Ande al livello del mare, è necessario che almeno 35 chilometri di crosta vengano asportati dall'erosione che finirà quindi con l'esporre la parte più profonda della crosta terrestre. I detriti sedimentari rimossi dall'edificio montuoso verranno accumulati sulla piattaforma continentale, sulla scarpata continentale e sul piano abissale del Pacifico. Là, i sedimenti si accumuleranno in attesa del prossimo ciclo tettonico che li spiegazzerà, li solleverà e li dislocherà per lasciare il posto alla prossima grande dorsale vulcanica.

La corona solare

La natura fisica dell'atmosfera più esterna del Sole emerge dalle osservazioni fatte non solo da terra durante le eclissi, ma anche dai razzi, dai satelliti e persino dalla Luna

di Jay M. Pasachoff

Il brillante disco visibile del Sole, la fotosfera, oscura alcune delle più interessanti caratteristiche solari. Durante un'eclisse totale di Sole, come quella particolarmente lunga dell'estate scorsa, la fotosfera viene completamente nascosta e si possono vedere alcune di queste caratteristiche. Una di essa è la cromosfera, lo strato irregolare dell'atmosfera solare immediatamente sovrastante la fotosfera. Durante un'eclisse di Sole, la cromosfera è visibile per pochi secondi all'inizio e alla fine della totalità come un chiarore rossastro lungo il bordo della Luna che avanza. Il colore rosso deriva dal fatto che la cromosfera ha un'intensa emissione a una certa lunghezza d'onda dello spettro dell'idrogeno nota come idrogeno-alfa. Quando anche il chiarore della cromosfera viene oscurato dalla Luna, allora emerge il pallido splendore della parte più esterna dell'atmosfera solare: la corona. Fino a questo secolo la corona era tanto misteriosa quanto imponente. Negli ultimi decenni se ne è gradualmente scoperta la natura mediante i telescopi e i calcolatori a terra e gli strumenti montati su razzi, satelliti e persino sulla Luna.

La corona è un milione di volte più debole della fotosfera. Di solito è nascosta, per un osservatore terrestre, dal normale cielo azzurro. In principio si dovrebbe poterla vedere semplicemente coprendo il disco solare col pollice. In pratica la luce diffusa del cielo in prossimità del Sole è ancora troppo intensa per permetterci di vedere la corona. Gli astronomi hanno però la fortuna di avere un pollice lunare. La Luna è 400 volte più piccola del Sole, ma è anche 400 volte più vicina alla Terra: ha perciò quasi esattamente lo stesso diametro angolare (nel cielo) della fotosfera solare, circa mezzo grado.

L'orbita lunare giace all'incirca sul piano dell'eclittica (il piano dell'orbita terrestre intorno al Sole). Ogni mese si avvicina al Sole nel cielo; la faccia più lontana dalla Terra è illuminata e quella vicina buia: in questa posizione prende il nome di luna nuova. L'orbita lunare è leggermente inclinata rispetto all'eclittica di modo che la Luna passa di solito sopra o sotto il Sole. Da due a cinque volte all'anno però la Luna passa di fronte al Sole. Talvolta l'ombra conica della Luna è leggermente spostata rispetto alla Terra e tutto quello che si può vedere è un'eclisse parziale. Altre volte la Luna si trova su un punto relativamente lontano dalla sua orbita intorno alla Terra e ha perciò un diametro angolare inferiore a quello del Sole. Durante una di queste eclissi un osservatore vede un anello di luce solare intorno alla Luna scura; si dice che l'eclisse è anulare.

Finché si vede anche solo una piccolissima frazione di fotosfera c'è tanta di quella luce diffusa dall'atmosfera terrestre che la corona rimane nascosta. Quando la Luna è però abbastanza vicina e opportunamente centrata sul disco del Sole, l'ombra lunare si sposta sulla Terra in una fascia larga 270 chilometri e lunga migliaia di chilometri. Da tutti i punti di questa fascia di totalità è visibile la corona per un periodo che può variare da una frazione di secondo al massimo teorico di sette minuti e 40 secondi, a seconda delle condizioni della singola eclisse e della posizione dell'osservatore lungo tale fascia.

Gli astronomi hanno dovuto spostarsi lungo tutto il globo per portare le loro apparecchiature in una località favorevole entro la fascia della totalità. Un cattivo funzionamento di uno strumento nel momento cruciale può rovinare i risultati di mesi di preparazione. Il tempo è sempre il fattore principale; molte spedizioni sono state guastate dalla pioggia o dal cielo coperto o anche da una nube che ha coperto il Sole nell'istante sbagliato. Le eclissi hanno attratto spedizioni recentemente nel Messico meridionale il 7 marzo 1970, in Canadà e in Alaska il 10 luglio 1972 e in Africa il 30 giugno del 1973.

La corona che si vede durante le eclissi ha tre principali componenti. La prima, più chiaramente visibile, è la corona interna, e si estende per circa due raggi solari dal centro del Sole, circa 1,4 milioni di chilometri. È resa visibile dalla luce fotosferica diffusa dagli elettroni nel gas che circonda il Sole. Il suo spettro è continuo, cioè non presenta quelle righe scure di assorbimento dette righe di Fraunhofer che rappresentano le lunghezze d'onda della luce assorbita entro la

La polarizzazione della corona è illustrata nella pagina a fronte in una immagine a colori sintetizzata dal calcolatore. Donald H. Menzel dell'Harvard College Observatory e l'autore hanno fotografato la corona attraverso filtri polarizzatori durante l'eclisse totale di Sole del 7 marzo 1970 a Miahnatlán in Messico. Le fotografie furono analizzate fotoelettricamente e l'informazione riguardo alla direzione e all'entità della polarizzazione di ciascuna fu accumulata nella memoria di un calcolatore. Lynn H. Quam e Robert B. Tucher dello Stanford Artificial Intelligence Laboratory usarono il calcolatore per combinare l'informazione delle fotografie e per generare un «display» a tre colori su uno schermo televisivo. Il colore corrispondeva alla direzione della polarizzazione e la saturazione al grado di polarizzazione. La corona è polarizzata radialmente in direzione esterna alla superficie del Sole. Sull'immagine di sintesi ciò è indicato dal fatto che i colori della corona sono complementari a intervalli di 90 gradi.

fotosfera. Questo componente della corona si chiama corona *K* dal tedesco *Kontinuum*.

A circa due raggi solari la corona *K* è meno dominante e a essa si somma la radiazione della luce solare diffusa dalla polvere che si trova distribuita nello spazio interplanetario. Lo spettro di questa seconda componente della corona presenta le righe scure di Fraunhofer e si chiama perciò corona *F*. La corona *K* continua nello spazio insieme alla corona *F* sebbene diventi piú debole al crescere della distanza dal Sole. Si estende anche al di là dell'orbita terrestre sebbene a queste distanze sia naturalmente assai rarefatta.

La terza componente principale della luce coronale è rivelata dalle righe di emissione nello spettro degli atomi altamente ionizzati vicino al Sole: atomi che hànno perduto fino a metà degli elettroni che a loro appartengono. Sebbene questa radiazione sia molto debole e porti un contributo trascurabile all'intensità totale della corona, fornisce precise informazioni sulle condizioni fisiche esistenti nella corona.

La corona ha una temperatura di circa due milioni di gradi Kelvin. Molti argomenti di carattere fisico portano ora a questo risultato sebbene il problema di come la corona possa avere una temperatura tanto alta abbia costituito per molti anni uno degli interrogativi principali della fisica solare. Lo spettro continuo è una traccia importante. I due milioni di gradi si riferiscono alla temperatura cinetica: indicano cioè che gli elettroni si muovono in modo cosí veloce che le lunghezze d'onda della luce fotosferica, ciascuna delle quali diffonde, sono spostate per effetto Doppler. Gli spostamenti sono assai grandi a causa dell'alta temperatura e sono completamente casuali in modo da eliminare tutte le righe di assorbimento dello spettro fotosferico di Fraunhofer. Storicamente è stato questo il primo ragionamento per spiegare la natura continua dello spettro della corona *K*. Solo quando le righe di emissione dello spettro coronale vennero identificate con la radiazione di atomi altamente ionizzati, gli astronomi solari si convinsero che la temperatura coronale era veramente alta.

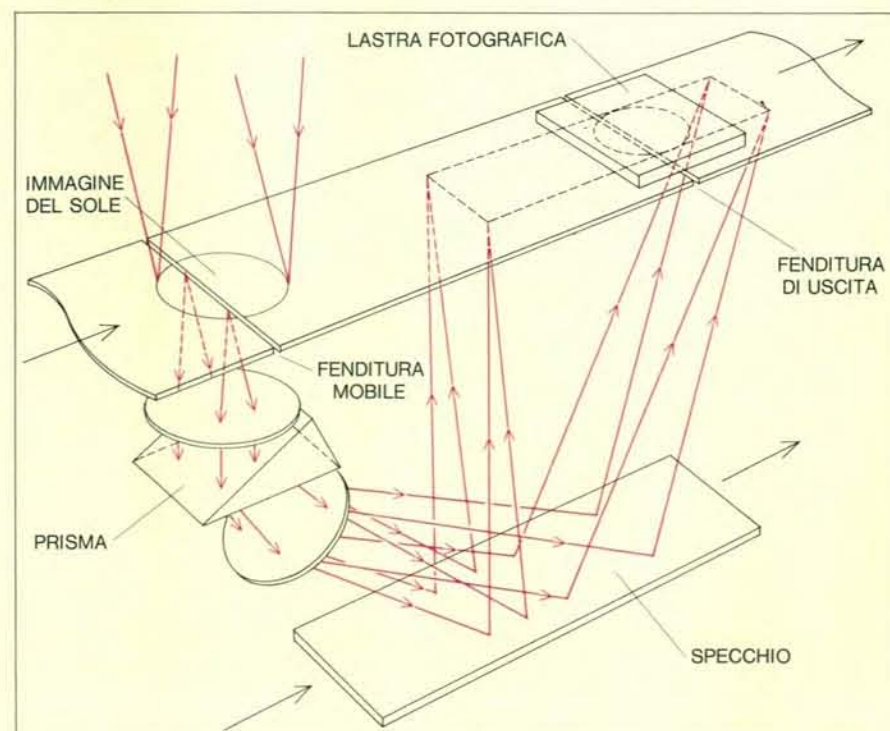
Si sono trovati sul Sole piú di 60 elementi chimici e si sa ora che la loro abbondanza nella corona è essenzialmente la stessa di quella esistente nella fotosfera. Due delle piú forti righe della regione visibile dello spettro sono la riga verde alla lunghezza d'onda di 5303 angstrom e la riga rossa a 6374 angstrom. La prima riga è emessa da atomi di ferro privati di 13 dei loro 26 elettroni dall'altissima temperatura, cioè da atomi di Fe XIV. La seconda riga è dovuta ad atomi di ferro privati di nove elettroni cioè atomi di Fe X. (Si indica lo stato non ionizzato di un elemento col numero romano I, lo stato una volta ionizzato con II e cosí via). Esistono due righe di emissione piú forti nel vicino infrarosso: le righe del Fe XIII a 10 747 e a 10 798 angstrom. Queste righe però si trovano in una regione spettrale relativamente poco esplorata in cui l'emulsione fotografica non è sensibile. Durante l'eclisse di giugno io e J. Philip Schierer della Tetronix Incorporation usammo un nuovo spettrometro elettronico per osservare queste righe. Sebbene i miei risultati siano ancora preliminari, mettono tuttavia in luce l'importanza dell'uso di dispositivi come i Vidicon a diodi al silicio per il vicino infrarosso.

Per liberare molti elettroni dal loro legame atomico il gas deve essere veramente caldo. Anche le righe di emissione indicano la temperatura elettronica in un altro modo: sono infat-

La struttura del Sole e la corona sono riassunti in questo diagramma in sezione. L'energia proveniente dalle reazioni termonucleari nel nucleo si sposta gradualmente verso gli strati esterni del Sole. La maggior parte della luce visibile ricevuta sulla Terra proviene dalla fotosfera; immediatamente sotto la fotosfera vi è la zona convettiva. Le onde d'urto provenienti dalla zona convettiva trasportano l'energia verso l'esterno nella cromosfera e nella corona. La temperatura del Sole ha il valore piú basso nella fotosfera e nella bassa cromosfera; le onde d'urto fanno sí che la temperatura aumenti attraverso l'alta cromosfera e la corona fino a raggiungere circa due milioni di gradi Kelvin nella corona. Dalla corona si estende il vento solare nello spazio interplanetario. La corona visibile ha tre componenti principali. La prima è la corona *K*: luce diffusa dagli elettroni nel gas che circonda il Sole. La seconda è la corona *F*: luce diffusa dalla polvere interplanetaria esistente tra il Sole e la Terra. La radiazione della terza componente è emessa da atomi altamente ionizzati vicino al Sole.

L'intera corona è stata fotografata durante l'eclisse totale del 30 giugno 1973 a Loiyengalani, in Kenia, da John W. Firor, Joseph H. Rush, Carl G. Lilliequist, Leon B. Lacey e Howard K. Hull dell'High Altitude Observatory di Boulder, Colorado. La fotografia è stata fatta in luce rossa alla lunghezza d'onda di 6500 angstrom attraverso un filtro di densità neutra la cui densità diminuisce radialmente dal centro verso l'esterno. Questi filtri nascondono la corona interna brillante per mettere in evidenza i deboli getti esterni. Si rendono cosí visibili i complicati dettagli della corona. L'alone circolare vicino all'estremità dei getti è dovuto al fatto che il filtro non ha celato uniformemente la luce del cielo particolarmente brillante durante l'eclisse.

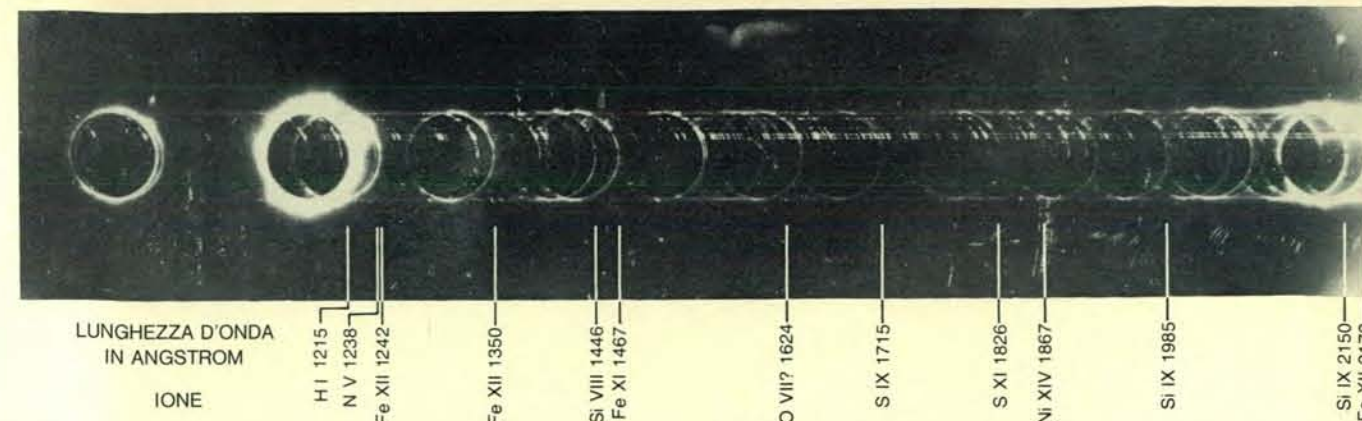
La corona interna è stata fotografata l'8 marzo 1970, il giorno dopo l'eclisse totale, con un coronografo al Mees Solar Laboratory dell'Università delle Hawaii sul bordo del cratere Haleakala a Mani. La fotografia è in realtà una composizione di sei immagini stampate insieme per ridurre la « grana » di una singola immagine. La fotografia è stata fatta in luce della riga in emissione « proibita » del ferro ionizzato 13 volte.

Il coronografo permette di fotografare e analizzare fuori eclisse la corona interna del Sole. La luce solare viene focalizzata su un cono occultatore che nasconde la fotosfera brillante. Un diaframma elimina un anello brillante di luce diffratta dal contorno della pupilla di entrata; una macchia nera al centro della seconda lente obiettiva elimina la macchia brillante che si forma al centro dell'immagine per riflessioni interne entro la lente principale. Una seconda lente obiettiva focalizza l'immagine della corona su un filtro, che isola la regione spettrale da fotografare.

Lo spettroeliografo fotografa il Sole a una particolare lunghezza d'onda. La luce solare penetra nella fenditura mobile e viene dispersa nello spettro di lunghezze d'onda mediante un prisma o un reticolo di diffrazione. Una lunghezza d'onda è isolata da una seconda fenditura. Via via che la prima fenditura spazzola il Sole per ricostruirne una immagine, la seconda fenditura si muove lungo la lastra fotografica con lo stesso moto in modo da ricostruire l'immagine in una stretta regione spettrale. Nel caso di spettroeliografi montati su satelliti Orbiting Solar Observatory (OSO) non si ottengono direttamente fotografie come quella della copertina di questo numero di « Le Scienze ». Le immagini vengono ricostruite da misure di splendore fatte fotoelettricamente, registrate su nastro magnetico e telemetrate a terra.

ti molto piú larghe delle righe di assorbimento della fotosfera che si formano a una temperatura media di soli 6000 gradi. Qui righe piú larghe significano temperature piú alte. Tutte queste righe sono « proibite » in condizioni ordinarie sulla Terra e ciò significa che la probabilità che la radiazione sia emessa a queste lunghezze d'onda è bassissima. Le righe solari proibite si hanno quando in un atomo un elettrone cade da uno stato di energia superiore che è metastabile, o di lunga vita, a uno stato di energia inferiore. La densità della corona è cosí bassa che gli atomi non sono diseccitati per collisione prima di aver avuto l'opportunità di irradiare energia.

La radioastronomia dà un quadro indipendente della corona. Via via che si esamina il Sole a lunghezze d'onda progressivamente piú lunghe si osservano strati sempre piú alti. Cosí a lunghezze d'onda centimetriche, la radiazione è emessa dalla cromosfera e alle lunghezze d'onda metriche è emessa dalla corona. In effetti quando si esamina il Sole a lunghezze d'onda via via piú lunghe, le dimensioni dell'immagine aumentano. Si è studiata la struttura spaziale della corona secondo per secondo con un radioeliografo composto di 96 antenne collegate tra loro, ciascuna di quasi 14 metri di diametro, a Cungoora in Australia. Si può studiare la struttura della corona anche mediante il radar: ricevendo l'eco di onde radio di diverse lunghezze d'onda. Queste osservazioni insieme ai calcoli teorici sono concordi nel dimostrare che la corona ha una temperatura di due milioni di gradi Kelvin.

L'atmosfera terrestre non lascia passare onde elettromagnetiche che siano piú corte di 3000 angstrom, corrispondenti cioè alla regione ultravioletta dello spettro. Si sono perciò inviati strumenti fuori dell'atmosfera per raccogliere informazioni sulla corona a lunghezze d'onda corte. Ne è stato pioniere Richard Tousey del Naval Research Laboratory che ha fatto per molti anni spettri del Sole con apparecchiature montate su razzi. Nella regione spettrale delle piú brevi lunghezze d'onda ultraviolette e alle lunghezze d'onda dei raggi X si possono vedere persino le righe spettrali « permesse » della corona e non solo quelle proibite.

Piú di recente i satelliti in orbita intorno alla Terra hanno fornito piattaforme stabili per osservare la radiazione ultravioletta della cromosfera e della corona. Leo Goldberg, che era allora all'Harvard College Observatory, guidò un gruppo che fece volare uno

Lo spettro della corona eseguito con uno spettrografo senza fenditura a bordo di un razzo lanciato da Wallops Islands, durante l'eclisse totale di Sole del 7 marzo 1970, mostrò nell'ultravioletto 25 righe spettrali precedentemente sconosciute. Tredici di esse sono illustrate in questa fotografia. Lo spettrografo non è dotato di fenditura perché l'immagine della corona o della cromosfera è cosí sottile da agire essa stessa da fenditura. Le righe spettrali cromosferiche sono brevi archi perché provengono da un bordo solare per volta. Le righe coronali sono emesse da strati piú alti sopra la superficie del Sole e sono rappresentate da cerchi completi. Un risultato particolarmente sorprendente di questo esperimento è stata la scoperta che nella corona esiste idrogeno non ionizzato. Sullo spettro si vede indicato dall'anello brillante identificato con HI. L'esperimento fu un tentativo congiunto dell'Harvard College Observatory, dell'Imperial College of Science and Technology dell'Università di York e dell'Astrophysics Research Unit del British Science Research Council presso il Culham Laboratory. Un elemento non ionizzato si indica col numero romano I; con II nel primo stato di ionizzazione e cosí via. H è idrogeno, N è azoto, Fe è ferro, Si è silicio, O è ossigeno, S è zolfo e Ni è nichel. Il numero che segue l'identificazione è la lunghezza d'onda della riga espressa in angstrom.

strumento il quale poteva funzionare sia da spettrografo che da spettroeliografo su due degli Orbiting Solar Observatory lanciati dalla NASA: OSO IV e OSO VI. Nella versione a spettrografo lo strumento analizzava lo spettro del Sole nel lontano ultravioletto. Si manteneva ferma l'immagine solare mentre si variava l'angolo di un reticolo di diffrazione.

Il reticolo disperdeva la radiazione solare nelle sue varie componenti. In questo modo si possono far penetrare le diverse lunghezze d'onda in un tubo fotomoltiplicatore che poteva perciò registrare l'intensità della radiazione solare a ciascuna lunghezza d'onda. Nella versione a spettroeliografo lo strumento lavorava in modo opposto. Gli spettroeliogrammi erano fatti mantenendo il reticolo di diffrazione a un certo angolo in modo che sul fotomoltiplicatore cadesse solo una lunghezza d'onda. Si esplorava l'immagine solare su uno schermo televisivo per ricostruire l'immagine del Sole a quella lunghezza d'onda.

La risoluzione dell'apparecchiatura di OSO IV era di un minuto di arco e fu migliorata a mezzo minuto di arco su OSO VI. Sebbene questa risoluzione non possa competere con la risoluzione di un secondo di arco raggiungibile a volte con strumenti terrestri, si sono però raccolte importanti informazioni nello studio della regione dell'estremo ultravioletto.

Tutte le regioni della corona con una certa temperatura e densità sono caratterizzate dall'emissione di un particolare ione. Una sequenza di spettroeliogrammi eseguiti a lunghezze d'onda corrispondenti a stati ionizzati sempre piú alti mostra che piú alti gradi di ionizzazione corrispondono ad altezze maggiori entro la corona. È cosí possibile ricostruire un modello tridimensionale della corona.

Durante l'eclisse del marzo 1970 fu eseguito con successo un nuovo esperimento su un razzo in volo. Era stato preparato in collaborazione dall'Harvard College Observatory, dall'Imperial College of Science and Technology dell'Università di York e dalla Astrophysic Research Unit del British Science Research Council presso il Culham Laboratory. Questo gruppo riuscí a ottenere spettri della corona e della cromosfera nella regione ultravioletta da 977 a 2200 angstrom e scoprí 25 nuove righe coronali. Siccome si tratta di una regione dello spettro preclusa agli osservatori da terra, non sono state ancora analizzate molte righe spettrali.

Un risultato supplementare del volo del razzo nel marzo 1970 fu la sorprendente scoperta che nella corona è visibile idrogeno non ionizzato. Fu rivelato nella radiazione della riga spettrale ultravioletta indicata con Lyman alfa, che è emessa da un atomo di idrogeno quando un elettrone salta al livello fondamentale dal livello di energia immediatamente superiore. La temperatura della corona è cosí alta che praticamente tutto l'idrogeno deve essere ionizzato. Ciò nonostante, la piccolissima frazione che rimane non ionizzata – forse un atomo su 10 – può diffondere abbastanza radiazione da essere rivelata.

Un altro campo di ricerche recenti è stato lo studio del Sole alle lunghezze d'onda dei raggi X tra uno e 100 angstrom, soprattutto per mezzo di razzi e satelliti da parte di un gruppo diretto da Riccardo Giacconi dell'American Science and Engineering di Cambridge. Dato che i raggi X sono troppo energetici e penetranti per essere focalizzati mediante l'ordinaria ottica dei telescopi, Giacconi e il suo gruppo hanno progettato un'ottica in cui i raggi sono riflessi a incidenza radente e con questo apparecchio sono riusciti a risolvere dettagli fino a un secondo di arco. I raggi X sono emessi soprattutto dalle « macchie calde » della corona situate molto al di sopra delle regioni attive della superficie solare. In queste macchie calde non è rara una temperatura di quattro milioni di gradi Kelvin, sebbene parte della radiazione non sia solo termica e possa riflettere le interazioni di particelle cariche col campo magnetico coronale.

In accordo con quanto si deduce dalle regioni visibile e radio dello spettro, le informazioni provenienti dalla regione ultravioletta e dai raggi X indicano chiaramente che la temperatura della corona è di qualche milione di gradi. Non si deve però dimenticare che questa temperatura non implica che sia presente una grande quantità di energia. La corona è infatti assai tenue. Alla base vi sono solo 10^9 elet-

troni per centimetro cubo, e il numero diminuisce rapidamente con l'altezza. Questa densità è 10 milioni di volte inferiore alla densità elettronica della fotosfera. Sebbene ogni singolo elettrone si muova ad alta velocità, e abbia perciò un'alta energia cinetica, il fatto che vi siano così pochi elettroni significa che l'energia totale della corona è decisamente bassa.

Nondimeno la domanda è: Da dove i singoli elettroni traggono la loro energia cinetica? L'energia si sprigiona dalla corona sotto forma di onde d'urto provenienti dai livelli inferiori del Sole, probabilmente da sotto la fotosfera nella zona convettiva (*si veda la figura a pag. 56*). La temperatura aumenta nella cromosfera e poi subisce un rapido salto a un milione di gradi in poche migliaia di chilometri immediatamente sopra la cromosfera.

Le righe di emissione degli atomi altamente ionizzati rendono possibile rivelare la corona anche al di fuori di un'eclisse solare. Si possono fare osservazioni da terra di queste righe in pochi luoghi ideali. In teoria le esigenze per tali osservazioni sono semplici: l'osservatore deve disporre di un cielo molto limpido e di un'ottica molto luminosa. In pratica è però difficile soddisfare a queste esigenze e si ebbe una grande conquista quando il francese Bernard Lyot inventò il coronografo nel 1930.

Un coronografo è di solito un cannocchiale rifrattore perché le microscopiche imperfezioni negli specchi di un telescopio riflettore diffondono troppa luce. Nel fuoco della lente primaria del cannocchiale si forma un'immagine del Sole su un piccolo cono occultante che ha esattamente le dimensioni dell'immagine della fotosfera. Bloccando l'intensa luce fotosferica si può sperare di scorgere la corona. L'immagine della corona è rifocalizzata e inviata a filtri che selezionano le lunghezze d'onda che si desidera osservare. Le osservazioni vengono fatte di solito attraverso filtri che isolano le lunghezze d'onda intorno a una riga di emissione coronale dato che

Lo spettro della corona sopra una regione quieta del Sole da 650 a 250 angstrom è stato misurato dallo spettrometro installato a bordo di OSO VI. Da 504 a 470 angstrom si vede lo spettro continuo dell'elio come un graduale aumento di intensità.

Lo spettro della corona sopra una regione attiva del Sole misurato da OSO VI in un'area di macchie solari e di brillamenti. L'intero spettro è circa dieci volte più intenso di quello della regione quieta della figura superiore. Anche certe righe spettrali, come la riga del magnesio nove volte ionizzato a 625 angstrom, sono rinforzate rispetto allo spettro di fondo.

queste righe sono più brillanti dello spettro continuo.

La costruzione di un buon coronografo è ancora qualcosa di simile a un'arte. I luoghi adatti per le osservazioni con il coronografo sono rari. Devono essere situati in alta montagna sopra gran parte dell'atmosfera terrestre e il cielo deve essere eccezionalmente libero da polveri.

Le isole Hawaii sono diventate siti di primo ordine per coronografi perché le condizioni meteorologiche generali mantengono il pulviscolo atmosferico sotto le alte montagne. L'Università delle Hawaii ha sviluppato un osservatorio in un luogo eccezionale a un'altezza di 3054 metri sul bordo del cratere Haleakala nell'isola di Mani. Qui John T. Jefferies, Jack B. Zirker e Frank Q. Orrall costituiscono un attivo gruppo di ricercatori coronali. Il cielo è tanto limpido che la diffusione della luce da parte dell'atmosfera vicina al disco solare è normalmente inferiore a 10 milionesimi dell'intensità totale del Sole (*si veda la figura alla pagina seguente*). Si possono fare osservazioni della corona per circa 275 giorni all'anno. L'High Altitude Observatory di Boulder nel Colorado ha esso pure un sito coronale nelle Hawaii sulla montagna di Mauna Loa nell'isola principale. Altri importanti siti coronografici nel mondo si trovano ad Arosa in Svizzera, sul Pic du Midi in Francia, a Kislovodsk in Russia e a Norikura in Giappone.

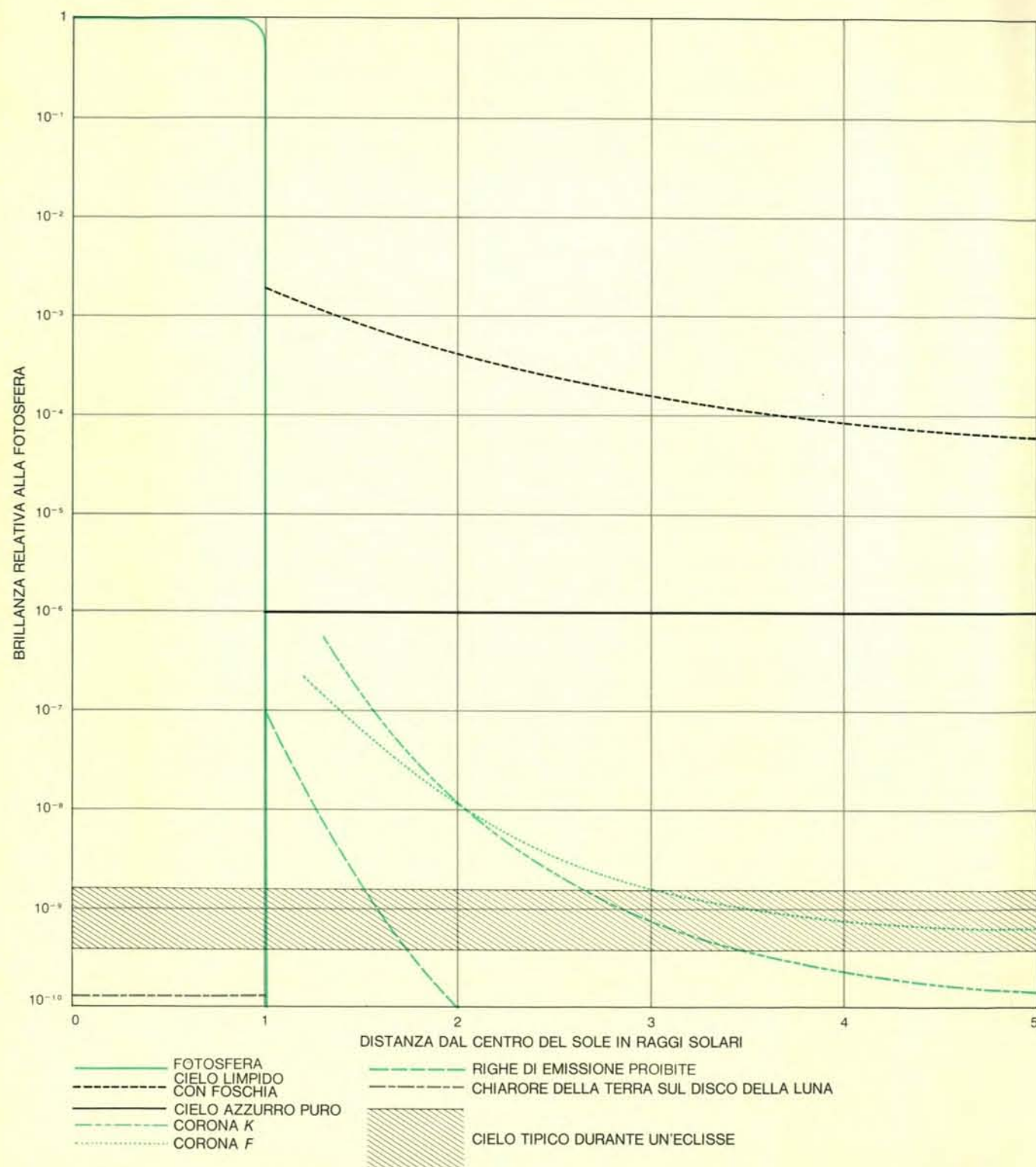
Un'immagine della corona in raggi X fu fatta da un razzo lanciato il 7 marzo 1970, giorno di un'eclisse totale di Sole. (Il razzo era fuori dell'area della totalità). Le lunghezze d'onda coprivano l'intervallo da 3 a 30 angstrom e da 44 a 55 angstrom. Le caratteristiche in raggi X corrispondevano a quelle osservate in luce bianca.

Al di fuori delle eclissi si può vedere solo molto difficilmente la corona *K*. Gli elettroni della corona *K* polarizzano la luce fotosferica e la diffondono. Si sono costruiti e si usano polarimetri in connessione coi coronografi per osservare la corona *K*. Queste osservazioni possono fornire informazioni sulle variazioni della corona interna da un giorno all'altro.

Alcune righe spettrali proibite di atomi altamente ionizzati sono state recentemente scoperte a lunghezze d'onda infrarosse maggiori di quelle che si possono osservare da terra. Durante l'eclisse solare del 12 novembre 1966 Guido Munch, Gerry Neugebauer e Dan McCammon degli Osservatori Hale hanno scoperto due nuove righe. Durante l'eclisse del marzo 1970, Kenneth H. Olsen, Charles R. Anderson e John N. Stewart del Los Alamos Scientific Laboratory hanno scoperto altre nove righe tutte nell'intervallo di lunghezze d'onda tra 10 000 e 30 000 angstrom. Tutte queste righe sono state osservate da un aereo che volava sopra la maggior parte del vapor acqueo terrestre che assorbe le lunghezze d'onda infrarosse. Gli aerei a reazione possono anche prolungare le eclissi inseguendo l'ombra della Luna; il supersonico *Concorde* prolungò la totalità dell'eclisse del 30 giugno 1973 fino al valore senza precedenti di 74 minuti. Si può studiare la relazione tra densità e struttura della corona e il suo campo magnetico sia da un punto di vista teorico che osservativo. Gordon Newkirk, jr. e Martin D. Altschuler dell'High Altitude Observatory hanno elaborato un metodo e un programma per calcolatori per derivare i campi magnetici nello spazio dalle regioni magnetiche osservate sulla superficie del Sole. I loro risultati possono essere rappresentati come diagrammi delle linee di forza magnetiche che si estendono nello spazio sopra la superficie solare. Le strutture coronali seguono le linee di forza magnetiche.

Kenneth Schatten e Norman F. Ness del Goddard Space Fligth Center della NASA e John M. Wilcox dell'Università della California a Berkeley hanno ideato un altro metodo per predire la struttura della corona. Per rendere più realistici i calcoli ammettono che la sorgente del campo magnetico interplanetario sia una superficie immaginaria di sei decimi di raggio solare sopra la fotosfera. Il campo magnetico si estende in direzione radiale dalla superficie e obbedisce a certe leggi semplici. Prima dell'eclisse del 22 settembre 1968 e del marzo 1970 Schatten fece delle previsioni su quella che avrebbe dovuto essere la struttura della corona e queste si manifestarono in ragionevole accordo con la struttura effettiva che infine apparve. Gerald W. Pneuman e Roger A. Kopp dell'High Altitude Observatory hanno elaborato una trattazione teorica più completa.

I telescopi solari che stanno ora volando nello spazio stanno rivoluzionando le osservazioni della corona. Su OSO VII, lanciato nel settembre 1971, era montato un coronografo con un disco occultante all'estremità di una lunga asta. I coronografi che girano nello spazio si possono costruire in maniera diversa dai loro analoghi usati a terra e possono raggiungere livelli mol-

to piú bassi di intensità. L'Apollo Telescope Mount dello Skylab, l'osservatorio satellite con uomini a bordo ora in orbita, monta anch'esso un coronografo. L'Apollo Telescope Mount ha schermi televisivi che forniscono in dettaglio la struttura della superficie del Sole in luce della riga idrogeno-alfa e alle lunghezze d'onda dei raggi X e della corona in luce bianca. Le esperienze principali dello Skylab comprendono uno spettroeliometro dell'Harvard College Observatory che fornisce una mappa del Sole a lunghezza d'onda tra 300 a 1350 angstrom con una risoluzione di cinque secondi di arco, un coronografo per luce visibile dell'High Altitude Observatory che dà fotografie che mostrano la struttura della corona tra 1,5 e 6 raggi solari a intervalli regolari, spettrografi ultravioletti nel Naval Research Laboratory e telescopi per raggi X dell'American Science and Engineering, della NASA e dell'Aerospace Corporation.

Intensità relative della corona, di altre parti del Sole e del cielo. Il diagramma si basa su quello fatto da H. C. van de Hulst. È evidente perché non si può osservare la corona con l'ordinaria luce diurna del cielo mentre essa appare invece chiaramente durante un'eclisse totale. La curva per le righe proibite riflette l'intensità di tutte queste righe nel Sole.

Adesso che ci sono informazioni intorno alla corona provenienti giornalmente da OSO VII e dallo Skylab, si possono confrontare le osservazioni con la teoria molto piú spesso di quello che sarebbe stato possibile se si fosse dovuto fare assegnamento esclusivamente sulle eclissi. Ci vorranno molti anni ancora prima che i fisici solari abbiano sintetizzato il gran numero di dati che stiamo ora ricevendo dallo spazio. Ciò nondimeno si può essere ragionevolmente sicuri che i modelli attuali della corona dovranno essere drasticamente riveduti.

Le osservazioni della corona forniscono sempre un'intensità totale che include non solo un contributo della corona *K*, ma anche quello della corona *F*. Si può determinare quanto contribuisca la corona *F*, perché differisce dalla corona *K* per due aspetti: per lo spettro e per la polarizzazione della luce. Io e Donald Menzel abbiamo eseguito una serie di osservazioni spettrografiche e polarimetriche durante le ultime quattro eclissi solari per studiare la separazione tra corona *K* e corona *F*. Mentre la luce diffusa dagli elettroni nella corona *K* è altamente polarizzata, la luce diffusa dalla fredda polvere interplanetaria della corona *F* è fondamentalmente non polarizzata. Il grado di polarizzazione della corona nel suo insieme raggiunge un massimo di circa il 40 per cento alla distanza di 1,5 raggi solari dal centro del disco solare. Se ne deduce che l'entità della polarizzazione diminuisce. Si può interpretare questa diminuzione come il risultato della sovrapposizione della corona non polarizzata *F* sulla corona polarizzata *K*. La stessa corona *K* raggiunge un livello di polarizzazione di circa il 60 per cento a circa tre raggi solari.

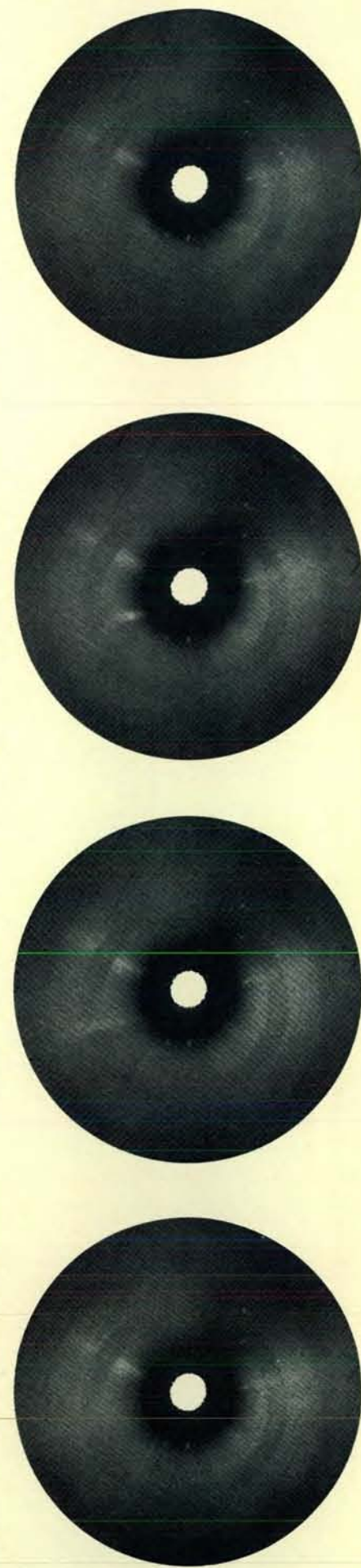
Per separare la componente polarizzata da quella non polarizzata, si deve registrare l'intensità della corona durante un'eclisse mediante polarizzatori posti ad angoli diversi. Il metodo ha però un limite fondamentale, fatto rilevare da D.R. Blackwell dell'Università di Oxford, e cioè che la luce della corona *F* non è completamente polarizzata, ma ha una polarizzazione residua di qualche per cento. Perciò i risultati che si ottengono per la separazione della corona *F* e della corona *K* al di là di circa cinque raggi solari, dove la polarizzazione totale è inferiore al 10 per cento, non sono precisi.

Per separare con precisione le due componenti si deve esaminare lo spettro della corona. Come già ricordato, la corona *F* diffonde semplicemente le righe di assorbimento di Fraunhofer della fotosfera solare e ciò non influisce sulla quantità assorbita dal livello fondamentale sottostante dello spettro continuo. Invece nella corona *K* tutte le righe sono scomparse. Per separare le due componenti coronali si deve misurare l'intensità delle righe di assorbimento in uno spettro fatto durante un'eclisse e confrontarla con le misure delle righe di uno spettro fotosferico fatto in un altro momento con lo stesso strumento. Il grado in cui si sono ridotte le intensità delle righe spettrali dà il contributo della corona *K*.

Questo metodo è stato usato per la riga idrogeno-alfa soprattutto da Blackwell e A.D. Petford. Ne hanno derivato un modello per le due componenti della corona che include la distribuzione della temperatura e della densità elettronica vicino al Sole fino alla distanza di 16 raggi solari. Molte loro osservazioni furono eseguite da un aereo a una quota di 9000 metri durante l'eclisse del luglio 1963. La tecnica è un perfezionamento del precedente lavoro di Blackwell durante l'eclisse del giugno 1954. Durante quel primo esperimento egli volò a 9000 metri in una carlinga aperta.

Una delle ragioni per cui la parte piú esterna della corona suscita tanto interesse è che c'è una grande lacuna tra la corona che si può osservare durante le eclissi e il debole chiarore all'orizzonte noto come luce zodiacale. La luce zodiacale è visibile in località prive di luci prima dell'alba e dopo il tramonto e si può osservarla fino a circa 80 raggi solari. Presumibilmente non è niente altro che un'estensione della corona *F*.

Una serie di dati che aiuta a colmare la lacuna proviene dai veicoli lunari senza uomini a bordo della serie Surveyor. Sulla superficie della Luna priva di aria non vi è ovviamente diffusione atmosferica della luce.

Un getto coronale si forma rapidamente nella corona sud-est e scompare rapidamente in un periodo di meno di cinque ore. È stato fotografato da un coronografo per luce visibile a bordo di OSO VII in quattro orbite successive il 9 maggio 1973. Il fotogramma superiore mostra il Sole alle ore 5 e 43 minuti del Tempo Medio di Greenwich prima che compaia il getto, il fotogramma successivo mostra l'inizio del getto alle ore 7 e 15 minuti; il terzo fotogramma mostra il getto, quasi scomparso, alle ore 8 e 47 minuti; il fotogramma in basso mostra il Sole alle ore 10 e 17 minuti dopo che il getto è scomparso. La scala di tempo dell'intero evento è molto piú breve di quanto si ritenesse in precedenza. Le fotografie sono state fornite dalla Rocket Spettroscopy Branch del Naval Research Laboratory di Washington.

La corona sorge semplicemente ogni mattina prima del disco solare e tramonta ogni sera dopo di esso. I. David Bohlin del Naval Research Laboratory analizzò alcuni dati del Surveyor 6 e del Surveyor 7. Poté seguire un getto coronale osservato dal Surveyor 7 fino a 22 raggi solari.

Anche le misure alle lunghezze d'onda radio possono venire utilizzate nella ricerca della densità elettronica della corona più esterna. La corona tende a spostare la posizione radio apparente della Nebulosa del Granchio o dei quasar quando questi le passano dietro. Inoltre cambia il periodo apparente dei segnali radio dei pulsar. Entrambi i tipi di variazioni dipendono dalla densità elettronica totale lungo la visuale.

La struttura della corona varia profondamente da un'eclisse all'altra. La sua forma è stata da lungo tempo associata al ciclo di attività solare legato al ciclo delle macchie solari di 11 anni. La forma della corona è piuttosto irregolare quando il numero delle macchie solari è al minimo (com'è ora); diventa molto più simmetrica quando il numero è al massimo. Al minimo delle macchie solari l'attività solare intensa è ristretta per la maggior parte alle regioni equatoriali del Sole e perciò in questo periodo la corona presenta soprattutto getti equatoriali. Al massimo dell'attività delle macchie l'attività solare è molto più generale e la corona si estende di più in tutte le direzioni.

L'occhio umano vede la corona fino a pochi raggi solari. I filtri a densità radiale, usati ora durante le eclissi, permettono di vedere un numero maggiore di dettagli. Un filtro a densità radiale è più scuro al centro che sui bordi e permette perciò di registrare le parti esterne meno luminose della corona entro la limitata latitudine di posa dell'emulsione senza sovraesporre l'immagine della parte interna più luminosa della corona. Queste fotografie mostrano con particolare nitidezza che la struttura appuntita dei getti coronali continua a esistere fino a grandi distanze dal Sole.

Questi getti costituiscono una prova che i modelli teorici della struttura della corona devono avere un certo numero di componenti. Un getto è probabilmente una regione in cui la densità elettronica è leggermente più alta che nelle regioni circostanti, di modo che diffonde più luce. I « buchi » scoperti recentemente nella corona sono regioni di densità molto inferiore.

Un getto ad anello (« loop ») di particelle ionizzate che scaturiscono dalla superficie del Sole e raggiungono fino a più di 40 000 chilometri nello spazio sono mostrate in questa immagine registrata elettronicamente e fatta con uno spettroeliometro ultravioletto a bordo del satellite Skylab. Lo strumento costituisce uno degli esperimenti di fisica solare dell'Apollo Telescope Mount ed è progettato da un gruppo dell'Harvard College Observatory guidato da Reeves. Lo spettroeliometro dello Skylab ha un potere risolutivo circa 30 volte maggiore di quello di strumenti simili in orbita sugli Orbiting Solar Observatory e rivela questi « loop » ben definiti alla lunghezza d'onda del ferro 14 volte ionizzato a 417 angstrom. Con misure ultraviolette si possono determinare in maniera molto precisa le temperature per livelli entro l'atmosfera solare. Queste misure sono fatte due volte al giorno durante i periodi della missione Skylab.

Ritengo che gli avvenimenti dell'anno passato si dimostreranno estremamente importanti per la nostra comprensione della corona. Quando si combineranno i risultati dell'eclisse di giugno ottenuti dallo Skylab e da OSO VII, cambieranno le nostre opinioni e miglioreranno le nostre teorie. Il quadro che sembra emergere è quello di una corona altamente dinamica con vaste regioni che cambiano spesso nel corso di minuti o ore.

Dovranno essere completamente riveduti i calcoli che si sono basati per molto tempo sull'ipotesi che la corona sia una nube di gas omogenea e simmetrica. Ci stiamo rendendo conto che non possiamo ignorare la complessità della corona, come quella della cromosfera, se vogliamo un modello del Sole che abbia un significato fisico e non sia soltanto un'astrazione teorica.

Questa è una lezione importante. Il Sole è solo la più vicina di miliardi di stelle, riguardo alla struttura delle quali ci accontentiamo spesso di concetti generali. Però sappiamo che altre stelle hanno cromosfere e corone e che molte di esse sono in uno stato più dinamico di quello del Sole. Le attività di queste stelle prendono parte all'attività costante che ora sembra sia la caratteristica di tanta parte dell'universo, che per lungo tempo è stato considerato in uno stato quanto mai tranquillo.

Memoria e apprendimento

Gli studi degli ultimi anni hanno portato a una critica radicale delle posizioni tradizionali e a una sempre più stretta interazione delle ricerche condotte nei due campi

di Francesco Robustelli

Memoria e apprendimento sono due processi intimamente connessi fra di loro. Per memoria possiamo intendere l'effetto durevole di una stimolazione, cioè la formazione e la conservazione di una traccia in seguito a una certa esperienza, e la possibilità di utilizzare questa traccia risalendo in qualche modo da essa all'esperienza che l'ha determinata. D'altronde, il fatto che un organismo vivente sia in grado di conservare gli effetti di una stimolazione dopo che essa è cessata, lo rende capace di modificare successivamente il proprio comportamento sulla base di quella stimolazione. Per apprendimento intendiamo appunto questa capacità. Occorre, tuttavia, tener presente che il comportamento può essere modificato da altri fattori oltre all'apprendimento, come ad esempio la fatica, l'età, le lesioni o la perdita di parti del corpo, la maturazione di modi di reazione innati, l'azione di farmaci.

Memoria e apprendimento sono anche, e di conseguenza, strettamente connessi fra di loro al livello metodologico. Non si può studiare l'una senza l'altro. Le misure della memoria non possono che essere inferite da una modificazione del comportamento, cioè dai risultati di un apprendimento. L'apprendimento, d'altronde, può essere capito solo sulla base delle informazioni che l'individuo ha ricevuto e conservato e per mezzo delle quali è riuscito a modificare il suo comportamento. Deriva da ciò il problema metodologico fondamentale degli studi sulla memoria: se non si è capito ciò che il soggetto ha appreso, non si potrà capire ciò che esso ricorda. (Capire ciò che il soggetto ha appreso significa fondamentalmente isolare nel complesso generale di una modificazione del comportamento le componenti dovute alla memoria, escludendo quindi quelle dovute alla fatica, alle lesioni, etc.). E viceversa, per quel che riguarda l'apprendimento: se non si sono prese in considerazione tutte le informazioni che il soggetto può aver ricevuto e conservato, non si potrà capire ciò che esso ha appreso. (La considerazione delle informazioni che il soggetto può aver ricevuto e conservato significa un'analisi accurata dei rapporti tra il soggetto e il suo ambiente, sia esterno che interno, nella situazione di apprendimento, secondo le possibilità proprie della specie a cui il soggetto appartiene ed eventualmente le particolari modalità dettate dalla sua storia individuale.) Il circolo vizioso è solo apparente, o almeno non piú reale di quanto non sia in qualsiasi altro campo di ricerca, dove sempre i vari processi inevitabilmente interagiscono fra di loro e quindi i metodi di indagine si integrano e si condizionano l'un l'altro, in modo piú o meno intricato.

In questo articolo mi propongo appunto di dare alcuni esempi significativi di come negli ultimi anni, da un lato, lo studio dei processi della memoria abbia portato a una nuova comprensione di certi aspetti dell'apprendimento e, dall'altro, lo studio dei processi dell'apprendimento abbia portato a una critica di certe nozioni tradizionali sui processi della memoria. Tutte le mie ricerche a cui farò riferimento in questo articolo sono state svolte in collaborazione con la dottoressa Nicoletta Del Grosso e la dottoressa Giovanna Spinozzi nei laboratori della Sezione di psicologia comparata dell'Istituto di psicologia del Consiglio nazionale delle ricerche a Roma, e in collaborazione con il dottor Murray E. Jarvik e la dottoressa Anne Geller nei laboratori del Dipartimento di farmacologia dell'Albert Einstein College of Medicine a New York.

Da almeno trent'anni la psicobiologia della memoria è concentrata su un problema e dominata da una teoria. È difficile dire se all'inizio sia stato scelto questo problema perché ritenuto particolarmente importante, e quindi siano state effettuate molte ricerche e presto ne sia derivata una teoria, oppure se proprio la disponibilità fin dall'inizio di una teoria indubbiamente assai plausibile abbia favorito lo studio di questo problema. Comunque, il problema è quello delle modalità con cui una esperienza, cioè l'effetto di un complesso di stimolazioni, viene fissata e conservata in qualche modo nel sistema nervoso di un organismo vivente. La teoria è quella del consolidamento della traccia mnemonica.

Questa teoria, in realtà, venne formulata per la prima volta nel 1900 da Müller e Pilzecker, i quali avevano notato che l'oblio di una prima lista di sillabe senza senso, determinato dall'apprendimento di una seconda lista, era funzione del periodo di tempo trascorso fra l'apprendimento delle due liste. Questi autori ipotizzarono che il verificarsi di un'esperienza desse l'avvio a processi nervosi « perseverativi », che continuavano cioè per un certo periodo dopo che l'esperienza si era verificata e cosí progressivamente ne consolidavano il ricordo. I loro risultati venivano quindi spiegati col fatto che l'attività nervosa provocata dall'apprendimento della seconda lista interferiva con i processi nervosi perseverativi relativi all'apprendimento della prima lista e che tale interferenza era tanto piú dannosa per il ricordo della prima lista quanto meno tempo questo aveva avuto per consolidarsi.

La teoria del consolidamento si mostrava particolarmente adatta a spiegare l'amnesia retrograda provocata da traumi cerebrali, in cui, come è noto, si ha un oblio selettivo degli eventi che si sono verificati subito prima del trauma, mentre il ricordo delle esperienze anteriori rimane intatto. Ma fu l'elettroshock a dare il massimo contributo alla teoria del consolidamento. Le frequenti amnesie dei pazienti sottoposti a questo trattamento, le quali regolarmente si estendevano su periodi piú o meno brevi precedenti la sua somministrazione, suggerirono immediatamente alcuni controlli sperimentali, che dimostrarono come i singoli eventi venissero dimenticati tanto piú facilmente quanto piú vicino alla somministrazione dell'elettroshock si erano verificati. Il passo successivo consistette, ovviamente, nell'inizio della sperimentazione con gli animali, che ha poi dato luogo a centinaia di ricerche nelle quali si sono potute liberamente manipolare tutte le variabili della situazione, come il tipo di apprendimento, l'intensità e la durata dell'elettroshock, l'intervallo fra l'apprendimento e la somministrazione dell'elettroshock, l'intervallo fra l'apprendimento e la prova di memoria, etc.

La teoria del consolidamento della traccia mnemonica ha avuto versioni differenti nei vari autori. Non è difficile, tuttavia, enuclearne l'assunzione fondamentale. Come è stato detto, la teoria cerca di spiegare in che modo il ricordo di un'esperienza venga fissato e conservato nel sistema nervoso, ma in effetti l'accento viene posto piú sulle caratteristiche temporali del processo di immagazzinamento dell'informazione che sulle modalità vere e proprie di tale processo (elettrochimiche, molecolari, etc.), che del resto differiscono nelle varie versioni della teoria. L'assunzione fondamentale, cioè, è in pratica rappresentata dal fatto che, una volta entrata nell'organismo un'informazione, il processo di immagazzinamento di tale informazione richiede un certo periodo di tempo. Durante questo periodo di tempo la traccia mnemonica (cosí viene chiamato l'ipotetico cambiamento, strutturale o funzionale, a carico del sistema nervoso, per mezzo del quale viene conservata l'informazione) « si consolida » progressivamente in seguito allo svolgersi del processo di immagazzinamento, e quindi è sempre piú resistente ai vari eventi che possono danneggiarla. Nel caso dei traumi cranici, quindi, la selettività dell'amnesia, che riguarda gli eventi che precedono piú da vicino il momento del trauma, si spiegherebbe col fatto che le tracce mnemoniche relative a questi eventi non avevano avuto il tempo di consolidarsi quando il trauma è avvenuto e perciò sono state cancellate o comunque danneggiate. Ma il risultato piú significativo in favore della teoria del consolidamento si ottiene in laboratorio somministrando un trattamento amnesico, per esempio l'elettroshock, a vari gruppi di animali, facendo variare l'intervallo di tempo che intercorre fra l'apprendimento e la somministrazione del trattamento. Poiché le tracce mnemoniche avrebbero tanto piú tempo per consolidarsi quanto maggiore è questo intervallo, la teoria prevede una graduazione dell'effetto amnesico del trattamento, che sarà massimo nel gruppo di soggetti al quale il trattamento è stato somministrato subito dopo l'apprendimento. In effetti, il risultato conferma ciò che la teoria prevede. Si ottiene quello che viene chiamato un « gradiente temporale di amnesia retrograda » (*si veda la figura di questa pagina*). Oltre che con l'elettroshock, tale gradiente può aversi con l'uso di altri trattamenti amnesici come ad esempio l'anossia, l'ipotermia, la *spreading depression*, la somministrazione di anestetici, di anidride carbonica, di inibitori della sintesi proteica (*si veda la figura a pagina 68*).

Con il moltiplicarsi delle ricerche, tuttavia, sono emersi due problemi fondamentali, che hanno messo sempre

Gradiente temporale di amnesia retrograda ottenuto con la somministrazione di un elettroshock. A gruppi differenti di soggetti è stato somministrato un elettroshock rispettivamente 10 minuti (*A*), 30 minuti (*B*), 6 ore (*D*) e 23 ore (*E*) dopo la prova di condizionamento; (*C*, gruppo di controllo). Nella figura sono indicate le prestazioni degli animali (topi) nella prova di memoria, effettuata una settimana dopo la prova di condizionamento. L'effetto amnesico dell'elettroshock varia in funzione inversa dell'intervallo fra prova di condizionamento e somministrazione dell'elettroshock.

Gradiente temporale di amnesia retrograda ottenuto con la somministrazione di un inibitore della sintesi proteica. A gruppi differenti di soggetti è stata somministrata la cicloesimide rispettivamente 1 minuto (*A*), 10 minuti (*B*) e 30 minuti (*D*) dopo la prova di condizionamento; (*C*, gruppo di controllo). Nella figura sono indicate le prestazioni dei topi nella prova di memoria, effettuata una settimana dopo la prova di condizionamento. L'effetto amnesico della cicloesimide varia in funzione inversa dell'intervallo fra prova di condizionamento e somministrazione della sostanza.

piú in imbarazzo i sostenitori della teoria del consolidamento.

Un problema riguarda la sconcertante diversità della lunghezza dei gradienti trovati dai vari autori. Per lunghezza del gradiente si intende il massimo intervallo di tempo fra apprendimento e somministrazione del trattamento che ancora consenta un effetto amnesico del trattamento stesso. La teoria prevede, infatti, che dopo l'apprendimento le tracce mnemoniche si consolidano sempre di piú fino al punto in cui diventano completamente resistenti all'azione di un determinato trattamento amnesico. La lunghezza del gradiente, quindi, dovrebbe corrispondere all'intero periodo di consolidamento della traccia mnemonica e pur dipendendo, ovviamente, da molte variabili, come il tipo di apprendimento, il tipo di trattamento amnesico, la specie animale, dovrebbe almeno fornire valori non molto lontani fra di loro, dato soprattutto che, nella maggioranza degli esperimenti, tali variabili sono assai simili.

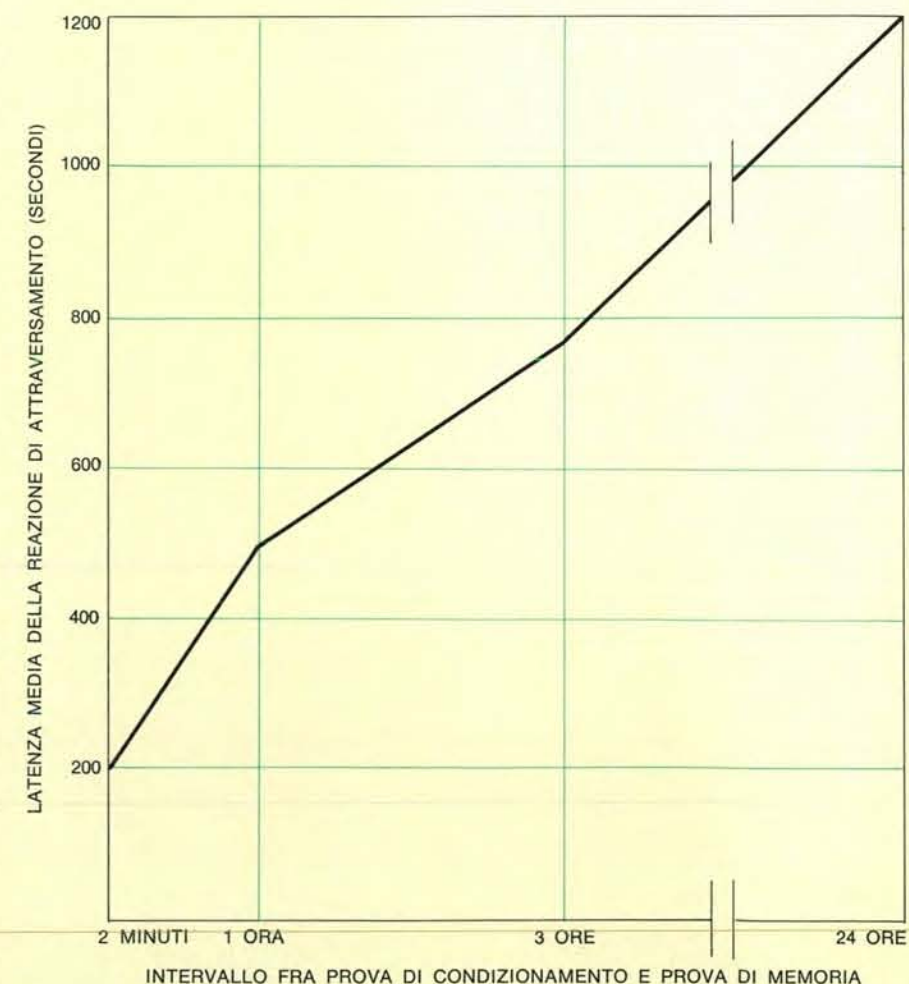
Purtroppo i gradienti trovati finora hanno lunghezze estremamente diverse. Il gradiente piú breve ottenuto con l'elettroshock, per esempio, è stato trovato nel 1965 da Chorover e Schiller e aveva una durata di dieci secondi. Il gradiente piú lungo è stato trovato da noi nel 1969 e aveva una durata di ventitrè ore. In altre parole, nell'esperimento di Chorover e Schiller la traccia mnemonica avrebbe impiegato dieci secondi per consolidarsi. Nel nostro esperimento avrebbe impiegato ventitré ore. Come si vede, si tratta di tempi neppure paragonabili fra di loro, data la diversità degli ordini di grandezza. Fra questi due estremi si collocano gli innumerevoli gradienti trovati dagli altri autori, alcuni della durata di secondi, altri di minuti, altri di ore.

L'altro problema è rappresentato dal fatto che in molti esperimenti l'elettroshock non ha provocato alcun effetto amnesico, anche quando sia stato somministrato subito dopo la fine della prova di apprendimento (si tenga presente che i procedimenti sperimentali permettono la somministrazione dell'elettroshock anche una frazione di secondo dopo la fine della prova di apprendimento). Ciò rende ancora piú difficile farsi un'idea di questo ipotetico processo di consolidamento della traccia mnemonica, che si completa in alcuni casi in una giornata e in altri casi in una frazione di secondo. La questione si complica ancora di piú se si prendono in considerazione i dati clinici, che pure la maggioranza dei sostenitori della teoria del consolidamento continua a considerare equivalenti ai dati sperimentali. Vengono riportati casi di pazienti sottoposti a trattamento con elettroshock in cui si sono avute amnesie retrograde permanenti che si estendevano su periodi di alcuni anni precedenti l'inizio del trattamento.

Nel 1960 veniva formulata da Coons e Miller la prima valida teoria alternativa a quella del consolidamento per spiegare l'effetto amnesico dell'elettroshock. Questi autori dimostrarono che l'elettroshock ha un effetto punitivo e che pertanto può essere considerato alla stessa stregua dei vari « rinforzi » normalmente usati nelle tecniche di condizionamento strumentale. Come è noto, in questo tipo di condizionamento l'uso opportuno di rinforzi positivi o negativi, cioè di ricompense o di punizioni, permette di selezionare determinate reazioni del soggetto nella particolare situazione in cui viene posto, nel senso che l'animale tende a ripetere le reazioni che sono state ricompensate (con cibo, acqua, etc.) e a inibire le reazioni che sono state punite (di solito con la somministrazione attraverso il pavimento dell'apparecchio di lievi shock elettrici, il cui unico scopo è quello di provocare del dolore). Nell'esperimento di Coons e Miller gli animali venivano puniti con un lieve shock elettrico se non passavano, entro un certo periodo di tempo, nell'altra parte dell'apparecchio. Subito dopo il passaggio, a un gruppo di soggetti veniva somministrato l'elettroshock e si otteneva il solito effetto amnesico. Cioè, il giorno dopo, questi soggetti non mostravano come gli altri una tendenza a passare nell'altra parte dell'apparecchio per evitare lo shock elettrico. Coons e Miller dimostrarono che questo avveniva non perché gli animali avessero dimenticato, in seguito alla somministrazione dell'elettroshock, ciò che avevano appreso, ma perché la somministrazione dell'elettroshock costituiva a sua volta una punizione, di modo che gli animali venivano a trovarsi in una situazione conflittuale. Se non passavano nell'altra parte dell'apparecchio venivano puniti con la somministrazione dello shock attraverso il pavimento. Se passavano, venivano puniti con la somministrazione dell'elettroshock (dato che l'elettroshock veniva di solito somministrato dopo la prova di apprendimento, e quindi in questo caso dopo il passaggio dei soggetti). Ne risultava un ritardo nel passaggio, che simulava uno stato amnesico. La teoria del « conflitto » di Coons e Miller, tuttavia, non poteva spiegare i casi in cui l'amnesia si riferiva a tipi di apprendimento in cui agli animali era invece richiesto di « non passare » in un'altra parte dell'apparecchio per evitare la punizione, poiché allora gli effetti punitivi dello shock elettrico e dell'elettroshock si sarebbero dovuti combinare e avrebbero dovuto portare di conseguenza a un rafforzamento della reazione appresa. Inoltre, è stato dimostrato che l'effetto condizionante dell'elettroshock, cioè la sua possibilità di agire come punizione, ha bisogno di alcune prove per manifestarsi e quindi non può spiegare l'amnesia ottenuta con un'unica prova di apprendimento seguita da un unico elettroshock.

Nel 1965 venne formulata da Lewis e Maher la teoria dell'« inibizione condizionata ». Essa si basa su un concetto di Pavlov secondo cui gli stati di inibizione possono proteggere l'organismo contro stimolazioni troppo intense e quindi pericolose. Se dunque il coma indotto dall'elettroshock viene considerato una reazione di inibizione protettiva contro l'intensa stimolazione costituita dall'elettroshock stesso, si può dire in termini di condizionamento pavloviano che il coma è la reazione incondizionata provocata dall'elettroshock. Allora, sempre secondo la logica del condizionamento pavloviano, stimoli che precedano la somministrazione dell'elettroshock (primo fra tutti l'apparecchio in cui avviene il condizionamento) possono trasformarsi in stimoli condizionati e quindi provocare come reazione condizionata uno stato di inibizione. Questo significa che il soggetto, rimesso nell'apparecchio per la prova di memoria, entra in uno stato di inibizione, che però questa volta non arriva al coma, ma secondo Lewis e Maher si manifesta in forme piú lievi, come la sonnolenza o una scarsa attività motoria. Di nuovo verrebbe simulata una situazione amnesica, mentre invece è lo stato di inibizione in cui il soggetto si trova che gli impedisce di comportarsi sulla base del precedente apprendimento.

La teoria del conflitto e la teoria dell'inibizione condizionata hanno messo in evidenza dei nuovi fenomeni di grande interesse e hanno notevolmente contribuito a migliorare i procedimenti seguiti negli esperimenti con l'elettroshock, ma non sono riuscite a prendere il posto della teoria del consolidamento della traccia mnemonica. Il loro merito maggiore è costituito dalla nuova prospettiva in cui esse hanno affrontato il problema dell'amnesia retrograda. Entrambe le ipotesi considerano l'elettroshock uno stimolo incondizionato. La differenza fra di esse sta nel fatto che tale stimolo incondizionato viene inteso dalla teoria del conflitto secondo lo schema del condizionamento strumentale, dalla teoria dell'inibizione condizionata secondo lo schema del condizionamento pavloviano. Le due teorie, dunque, hanno messo da parte il punto di vista tradizionale (proprio della teoria del consolidamento) secondo cui si ha prima una « esperienza di apprendimento » e successivamente la somministrazione di un « trattamento amnesico », dimostrando l'artificialità di questa dicotomia in molti casi. Secondo la nuova prospettiva, si ha nella realtà solo una complessa situazione stimolante costituita da una serie di elementi che agiscono in una certa successione e quello che è tradizionalmente considerato il « trattamento amnesico » non è altro che l'ultimo stimolo o l'ultimo gruppo di stimoli in questa successione. Il fenomeno dell'amnesia retrograda, insomma, non viene spiegato in termini di « memoria », ma in termini di « apprendimento ». La teoria del conflitto e la teoria dell'inibizione condizionata sono state, dunque, importanti nella storia recente degli studi sui processi della memoria in quanto hanno incrinato lo schema ormai cristallizzato, sia concettuale sia metodologico, imposto dalla teoria del consolidamento, rivendicando la necessità di un'analisi piú approfondita delle varie situazioni di apprendimento e delle interazioni di tipo integrativo fra esse e i cosiddetti trattamenti amnesici.

L'interesse per un'analisi accurata della situazione di apprendimento e la convinzione che da tale analisi pos-

Incubazione della reazione condizionata. La forza della reazione condizionata di evitamento di tipo passivo, corrispondente alla latenza della reazione di attraversamento, aumenta progressivamente dopo la prova di condizionamento. Nella figura sono indicate le prestazioni di quattro differenti gruppi di topi nella prova di memoria.

sa scaturire la spiegazione di fenomeni tradizionalmente attribuiti all'area della memoria caratterizzano anche le altre due principali teorie formulate successivamente per spiegare l'amnesia retrograda: la teoria degli « effetti proattivi » e quella dell'« incubazione della reazione condizionata ». Si tratta di teorie meno definite delle precedenti, ma indubbiamente piú importanti, poiché sono state esse a mettere in seria difficoltà la teoria del consolidamento.

Per teoria degli effeti proattivi si può intendere quell'insieme di contributi sia teorici sia sperimentali che hanno posto in primo piano l'importanza di quegli effetti, temporanei o permanenti, che un trattamento considerato amnesico può avere sul comportamento, ma che non sono dovuti a un'interferenza con i processi della memoria. Si tratta di alterazioni di carattere fisiologico, come variazioni dell'attività spontanea, minore capacità di inibizione dei movimenti, influenze varie sui meccanismi della fame e della sete, etc. Gli effetti proattivi hanno molta importanza perché possono facilmente simulare un'alterazione della memoria. Se, per esempio, l'animale ha imparato a non lasciare una certa parte dell'apparecchio perché le altre parti hanno il pavimento percorso da corrente elettrica, un'eventuale diminuzione della capacità di inibire i suoi movimenti può costringerlo a lasciare tale parte, comportandosi come se avesse dimenticato l'esperienza di apprendimento. Oppure, una diminuzione della fame in un test basato su questa motivazione può portare a un peggioramento delle prestazioni dei soggetti, che può essere erroneamente interpretato come stato amnesico.

Schneider e Sherman (1968) hanno dato un contributo particolarmente interessante a questa teoria. Questi autori hanno messo in evidenza il fatto che la somministrazione del lieve shock elettrico, usato come punizione nella maggioranza dei test di apprendimento animale, produce due effetti diversi. Uno è appunto quello punitivo, che determina il condizionamento, agendo come stimolo incondizionato o come rinforzo negativo di determinate reazioni dell'animale, che quindi vengono soppresse. L'altro effetto consiste in un'azione eccitante su tutto l'organismo dell'animale, il quale quindi presenta una serie di reazioni fisiologiche e biochimiche alla somministrazione dello shock elettrico in sé, indipendentemente dai processi di apprendimento e di memoria. Tali reazioni hanno ovviamente un certo decorso e variano qualitativamente e quantitativamente in funzione di molti fattori, primo fra tutti il numero e l'intensità degli shock ricevuti dagli animali. Dopo l'esperienza di apprendimento, quindi, si ha un certo periodo in cui l'organismo del soggetto subisce una serie di alterazioni di carattere fisiologico-biochimico che interagiscono con i vari trattamenti, modificando la reattività del soggetto e quindi l'entità degli effetti dei trattamenti stessi. La somministrazione di un elettroshock subito dopo una prova di apprendimento avrebbe un maggior effetto non perché agisce su una traccia mnemonica ancora debole, ma perché l'organismo molto eccitato in quel momento reagisce piú intensamente all'elettroshock, risultandone danneggiati temporaneamente o permanentemente meccanismi non relativi alla memoria, che però influiscono in modo determinante sulle prestazioni degli animali. Da qui l'importanza di considerare la situazione di apprendimento nella totalità dei suoi aspetti e delle sue conseguenze, invece di isolare artificialmente un quadro limitato di fenomeni con il risultato di attribuire alla memoria ciò che con la memoria non ha nulla a che fare. Il soggetto, cioè, non deve essere considerato in una unilaterale prospettiva psicologica che contempli solo la sua possibilità di formare tracce mnemoniche, ma come un organismo vivente in tutta la sua complessità fisiologica e biochimica.

Questa nuova prospettiva è ancora piú completa nella teoria dell'incubazione della reazione condizionata, in cui confluisce una serie di contributi teorici e sperimentali piuttosto differenti, ma tutti centrati sul problema delle curve dell'oblio. Si tratta di contributi che hanno messo in evidenza come in molti casi l'andamento temporale delle prestazioni dei soggetti, che dovrebbero variare in funzione del ricordo, è completamente diverso da quello che ci si dovrebbe aspettare sulla base del concetto tradizionale di oblio, inteso come progressiva diminuzione del ricordo col passare del tempo. Nel cosiddetto effetto Kamin, ad esempio, le prestazioni dei soggetti diminuiscono progressivamente dopo l'esperienza di apprendimento fino a un valore minimo che spesso corrisponde all'oblio totale, per poi aumentare di nuovo progressivamente. Nel fenomeno dell'incubazione della reazione condizionata si ha esattamente l'opposto dell'oblio tradizionale, cioè il progressivo aumento della forza della reazione condizionata (*si veda la figura a pagina 69*). Con un lungo esperimento noi abbiamo potuto dimostrare come anche nel caso in cui si verifica l'incubazione si ha poi un oblio vero e proprio. Cioè, si ha una prima fase di progressivo aumento della forza della reazione condizionata fino al raggiungimento di un valore massimo e quindi una progressiva diminuzione (*si veda la figura a pagina 70*).

Incubazione della reazione condizionata e oblio. La forza della reazione condizionata di evitamento di tipo passivo aumenta progressivamente dopo la prova di condizionamento, raggiunge un valore massimo, e quindi progressivamente diminuisce. Nella figura sono indicate le prestazioni di sei differenti gruppi di topi nella prova di memoria. Gli intervalli fra prove di apprendimento e di memoria sono rappresentati in scala logaritmica.

Il fenomeno dell'incubazione ha suggerito una nuova spiegazione dei gradienti temporali di amnesia retrograda. Un agente amnesico risulterebbe sempre meno efficace col passare del tempo dopo l'apprendimento per il fatto che col passare del tempo la reazione condizionata stessa diventa sempre piú forte. Va chiarito che per la teoria del consolidamento si verifica dopo l'apprendimento un progressivo rafforzamento della traccia mnemonica, cioè dell'ipotetico cambiamento al livello del sistema nervoso mediante il quale verrebbe fissata e conservata la informazione, senza che ciò implichi alcun cambiamento della forza con cui si manifesta la reazione condizionata. La teoria dell'incubazione, invece, pone l'accento sulle conseguenze dell'esperienza di apprendimento (quali la reazione allo stress costituito da tale esperienza, le variazioni dello stato emozionale, etc.) che influiscono sul comportamento del soggetto e quindi sulla forza con cui una reazione condizionata si manifesta, indipendentemente ed eventualmente anche in assenza di corrispondenti variazioni al livello della memoria. La teoria dell'incubazione insiste anche sull'importanza di conoscere il normale andamento temporale della forza della reazione condizionata per poter valutare le variazioni di questa forza in seguito alla somministrazione di un trattamento, e in particolare insiste sull'importanza di conoscere la forza della reazione condizionata nel momento in cui il trattamento viene somministrato.

Un ulteriore passo avanti è stato da noi compiuto recentemente sulla base della teoria dell'incubazione. Precedenti esperimenti avevano dimostrato che, con opportuni parametri della situazione sperimentale, si ottiene non una semplice curva di incubazione, ma un andamento bifasico della forza della reazione condizionata che, dopo l'apprendimento, dapprima aumenta, poi diminuisce, poi aumenta di nuovo (*si veda la figura in questa pagina*). Somministrando l'elettroshock, si vede in questo caso che il suo effetto amnesico dapprima diminuisce, poi aumenta, poi di nuovo diminuisce. Sparisce cosí il gradiente temporale di amnesia retrograda previsto dalla teoria del consolidamento della traccia mnemonica, mentre viene confermata una correlazione negativa fra la forza con cui la reazione condizionata si manifesta e l'effetto del trattamento amnesico. Solo future ricerche potranno chiarire il significato di questa correlazione, che però suggerisce che quegli stessi fattori, che sono responsabili della forza con cui la reazione condizionata si manifesta in un dato momento, determinano in qualche modo l'effetto del trattamento amnesico. Ma poiché è impossibile che questi fattori riguardino il consolidamento della traccia mnemonica (altrimenti non si potrebbero avere andamenti bifasici), essi devono essere identificati o in altri meccanismi della memoria diversi dal consolidamento della traccia mnemonica, o in conseguenze dell'esperienza di apprendimento non riferibili all'area della memoria, oppure infine in qualche aspetto dell'esperienza di apprendimento riferibile all'area della memoria, ma non messo in evidenza finora. Le ricerche degli ultimi due o tre anni indicano che queste tre direzioni di ricerca sono egualmente proficue. Certo è ancora troppo presto per arrivare a delle conclusioni generali, ma può essere utile un rapido sguardo ai risultati di queste ricerche.

Andamento bifasico della forza della reazione condizionata di evitamento di tipo passivo. Dopo l'apprendimento, la latenza della reazione di attraversamento dapprima aumenta, poi subisce una diminuzione e, successivamente, aumenta di nuovo. Nella figura sono indicate le prestazioni di sei differenti gruppi di topi nella prova di memoria.

Per quel che riguarda gli altri eventuali meccanismi della memoria con cui potrebbero interferire gli eventi o i trattamenti amnesici, si sta già chiaramente affermando una teoria del recupero. Gli autori che la sostengono ricordano che in nessun caso si può pensare che i processi della memoria si esauriscano nella formazione di una traccia mnemonica, qualunque sia il tempo necessario per questa formazione. In ogni caso uno stadio importante di questi processi non può non essere costituito dal recupero, dall'uso dell'informazione che è stata immagazzinata. Solo attraverso l'attivazione della traccia mnemonica, infatti, e l'uso dell'informazione che in essa è stata fissata e conservata si ha l'apprendimento, cioè una modificazione del comportamento.

Il punto di partenza di questa nuova teoria è la constatazione che in parecchi casi la reazione condizionata, dapprima scomparsa in seguito alla somministrazione del trattamento amnesico, ricompare poi spontaneamente o per ef-

fetto di determinate condizioni. Se, dunque, il ricordo dell'esperienza di apprendimento ricompare senza che nel frattempo si sia verificato alcun nuovo apprendimento, ciò può significare solamente che tale ricordo non era mai stato cancellato dalla somministrazione del trattamento amnesico, ma solo reso inutilizzabile per un certo periodo di tempo. L'amnesia retrograda, totale o parziale, prevista dalla teoria del consolidamento, deve essere definitiva. I risultati sperimentali addotti dai sostenitori della teoria del recupero suggeriscono la possibilità che il consolidamento della traccia mnemonica sia

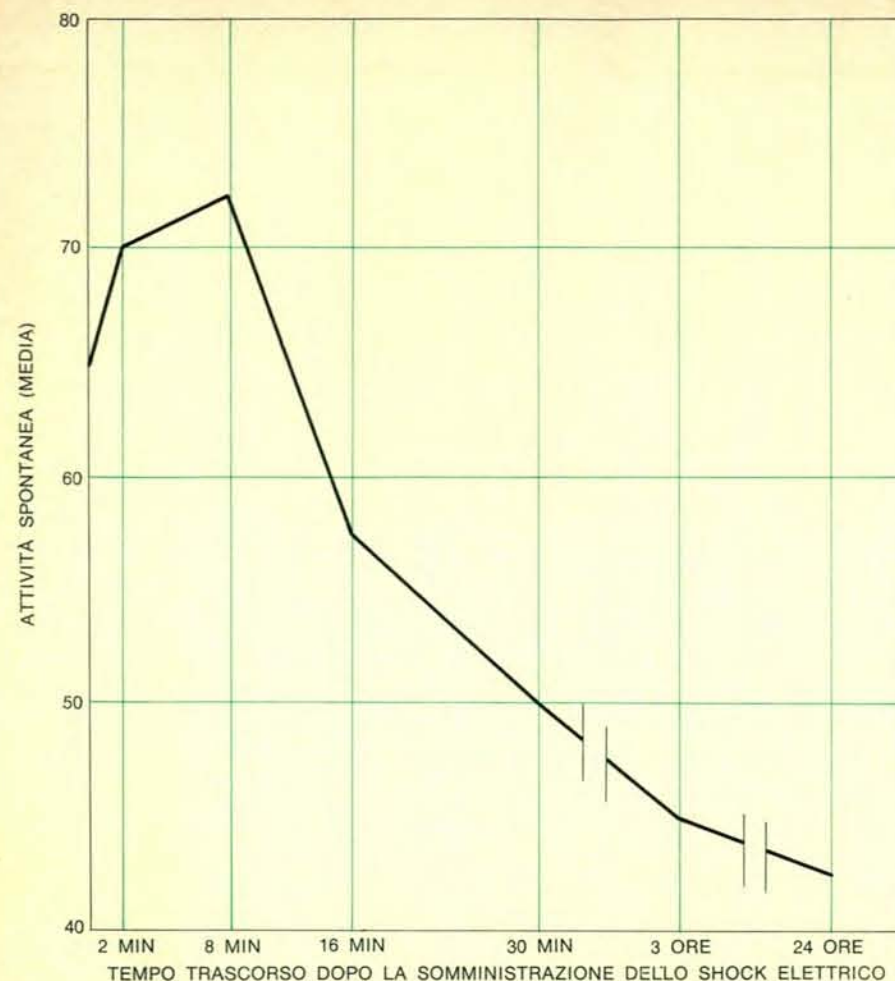
Condizionamento di evitamento di tipo passivo. L'animale è stato posto nella parte illuminata dell'apparecchio (*fotografia in alto*), la quale comunica per mezzo di una piccola apertura con una parte buia. Dopo un breve periodo di immobilità dovuta alla paura, l'animale entra nella parte buia per esplorarla. Qui riceve un lieve shock elettrico somministrato attraverso il pavimento, per cui fugge nella parte illuminata.

in realtà rapidissimo (una frazione di secondo) e che ciò che richiede tempo, invece, è la formazione di quel meccanismo, presumibilmente assai complesso, per mezzo del quale la informazione ritorna disponibile. Come spiegano Miller e Springer (1973), i due principali sostenitori di questa teoria, non è sufficiente che una biblioteca acquisti un libro e lo metta in uno scaffale. Mettere un libro in uno scaffale, essi dicono, è un'operazione analoga al consolidamento della traccia mnemonica. Catalogare il libro è l'operazione corrispondente alla formazione del meccanismo di recupero. Se questa seconda operazione non viene effettuata, il libro c'è in effetti nella biblioteca, ma non è possibile disporne quando se ne ha bisogno. Secondo la teoria del recupero, dunque, gli eventi e i trattamenti amnesici danneggerebbero il meccanismo di recupero, non quello di consolidamento della traccia mnemonica. Nessuna ipotesi, d'altronde, è stata formulata finora sull'andamento temporale dei processi che portano alla formazione del meccanismo di recupero. Miller e Springer fanno anche notare che, mentre la ricomparsa dell'informazione dopo l'amnesia prova che l'informazione stessa non poteva essere stata cancellata, ma era solo stata non disponibile per un certo periodo di tempo, i casi in cui l'informazione non ricompare non possono provare il contrario, in quanto si può sempre supporre o che non si siano trovate le condizioni adatte per sbloccare il meccanismo di recupero o che il danno arrecato a questo meccanismo sia irreversibile. Cioè, con le tecniche disponibili attualmente, non è mai possibile stabilire con certezza che una traccia mnemonica è stata cancellata ed escludere, quindi, la possibilità che essa sia intatta, ma non recuperabile. Il che equivale a dire che nessun tipo di risultato sperimentale, per ora, può confermare la teoria del consolidamento e smentire quella del recupero.

Per quel che riguarda le conseguenze dell'esperienza di apprendimento non riferibili all'area della memoria, continuano ad accumularsi i dati che indicano come semplici variazioni di carattere fisiologico possano modificare il comportamento del soggetto. Sono particolarmente numerose le ricerche che riguardano il condizionamento di evitamento di tipo passivo, di gran lunga il più usato per lo studio dei processi della memoria. Nella sua versione più comune questo tipo di condizionamento si realizza nel modo seguente (*si veda la figura in questa pagina*). L'animale viene posto nella parte illuminata di un apparecchio, la quale comunica per mezzo di una piccola apertura con una parte buia. Dopo un breve periodo di immobilità dovuta alla paura, per il fatto di trovarsi in un ambiente nuovo, l'animale passa nella parte buia dell'apparecchio per esplorarla, ma qui riceve un lieve shock elettrico attraverso il pavimento. L'animale fugge nella parte illuminata, da dove viene preso e rimesso nella gabbia in cui vive normalmente. Una cellula fotoelettrica posta all'interno della parte buia dell'apparecchio permette sia la somministrazione istantanea dello shock, quando il soggetto abbia raggiunto un punto convenzionale, sia il controllo di un timer che dà il tempo esatto trascorso dal momento in cui l'animale è stato posto nell'apparecchio al momento in cui ha raggiunto il punto convenzionale nella parte buia di questo. Dopo un certo intervallo di tempo, l'animale viene di nuovo rimesso nella parte illuminata dell'apparecchio. Se non ricorda nulla di ciò che è accaduto nella prima prova, tenderà a ripetere lo stesso comportamento della prima prova. Se al contrario ricorda, non passerà o passerà assai più tardi, nella parte buia dell'apparecchio, dove nella prima prova ha ricevuto lo shock elettrico. Il tempo impiegato dall'animale per passare da una parte all'altra dell'apparecchio viene chiamato « latenza » della reazione di attraversamento (per « reazione di attraversamento » si intende appunto il passaggio dell'animale « attraverso » l'apertura che mette in comunicazione le due parti dell'apparecchio). Il confronto fra la latenza della prima prova e quella della seconda ci dice se si è formata la reazione condizionata, oppure se essa è stata conservata. Il condizionamento di evitamento di tipo passivo (chiamato così perché il soggetto impara a evitare lo shock *non* passando nell'altra parte dell'apparecchio, cioè *non* dando una certa risposta motoria, mentre nel condizionamento di evitamento di tipo attivo il soggetto evita lo shock dando una certa risposta motoria) è il più usato nelle ricerche sui processi della memoria perché è di brevissima durata (10-20 secondi in genere) e si realizza in un'unica prova. Ciò permette, da un lato, di isolare un'unica esperienza relativamente semplice e di seguire lo sviluppo della corrispondente traccia mnemonica, mentre nei tipi di apprendimento che si realizzano in più prove c'è il problema, non risolvibile all'attuale stato delle nostre conoscenze, di come interagiscono fra di loro le varie prove ripetute. D'altro lato, la brevità dell'esperienza di apprendimento permette di calcolare con notevole precisione quanto tempo dopo questa esperienza venga somministrato un trattamento.

Il condizionamento di evitamento di tipo passivo si presta anche molto bene per studiare le conseguenze propriamente fisiologiche dell'esperienza di apprendimento. La latenza della reazione di attraversamento, infatti, come la misura di qualsiasi altro comportamento acquisito, non dipende solo dal ricordo dell'esperienza di apprendimento. Altri fattori possono modificarla. Ad esempio, una qualsiasi variazione dell'attività spontanea dell'animale porterà inevitabilmente a una variazione della latenza della reazione di attraversamento, indipendentemente da qualsiasi variazione del ricordo dell'esperienza di apprendimento. I risultati di un nostro esperimento hanno indicato che, dopo la semplice somministrazione di uno shock elettrico, dello stesso tipo di quelli usati per il condizionamento, l'attività spontanea di un topo dapprima aumenta e poi diminuisce gradatamente col passare del tempo (*si veda la figura a pagina 74*). Questo semplice fenomeno potrebbe spiegare l'incubazione della reazione condizionata, poiché una diminuzione dell'attività spontanea porta a un aumento della latenza della reazione di attraversamento, che di solito viene considerata una misura della memoria del soggetto. I risultati di un altro nostro esperimento hanno indicato che le stesse variazioni di attività spontanea (seppure di minore entità) si hanno dopo che il topo, invece di aver ricevuto uno shock elettrico, è stato per un minuto in un *open field*, cioè in un ampio spazio quadrato chiuso da quattro pareti verticali. Evidentemente, basta l'azione eccitante di un'esperienza nuova, anche in assenza di qualsiasi stimolazione nocicettiva, a determinare le suddette variazioni dell'attività spontanea. Se si pensa che ogni esperienza di apprendimento è per definizione un'esperienza nuova, si potrà avere un'idea di quanto semplici possano essere le cause di certe modificazioni del comportamento attribuite alla memoria e di come queste cause possano interagire con i vari trattamenti per determinare ulteriori modificazioni, anch'esse del tutto indipendenti dai processi della memoria.

Per quel che riguarda, infine, l'importanza che ha per lo studio dei processi della memoria l'esatta comprensione dell'esperienza di apprendimento anche a livello di ciò che è stato memorizzato, si può senz'altro affermare che i risultati in questo campo stanno a poco a poco cambiando alcuni dei nostri concetti fondamentali sul condizionamento. Si afferma sempre più l'idea, non nuova ma solo di recente sviluppata adeguatamente, secondo cui nella maggioranza dei casi la situazione di apprendimento è tale che non si formano solo le reazioni condizionate volute dallo sperimentatore, ma anche altre, sia per la presenza di parecchi stimoli in detta situazione, sia per il fatto che il rinforzo può agire casualmente su altre reazioni del soggetto oltre a quella che lo sperimentatore ha in mente. Ne deriva la necessità di stabilire che cosa esattamente il soggetto ha imparato o, se si preferisce, *tutto* quello che il soggetto ha imparato. È altrimenti impossibile capire il significato anche di una sola delle reazioni condizionate che si sono formate, dato che le varie reazioni condizionate interagiscono fra di loro e la prestazione che viene misurata è il risultato di tale interazione, oltre che (come si è più volte ripetuto) degli altri aspetti dell'esperienza di apprendimento. Soprattutto per l'influsso del behaviorismo americano, nel quadro del quale si è sviluppata la maggior parte delle tecniche per lo studio del comportamento animale, si è data una importanza preminente alle reazioni motorie, che spesso sono state arbitrariamente isolate come se fossero l'unico risultato delle esperienze di apprendimento, rinunciando a un'analisi completa di queste ultime, che sono sempre assai complesse, anche nei casi apparentemente più semplici.

Noi abbiamo svolto un gran numero di ricerche per analizzare il condizionamento di evitamento di tipo passivo, data l'importanza di questo tipo di condizionamento per il vasto uso che se ne fa nel campo degli studi sui processi della memoria. I risultati di queste ricerche indicano che tale condizionamento è in realtà un complesso misculio di condizionamenti sia classici sia strumentali. L'idea tradizionale secondo cui una specifica reazione motoria (nel nostro caso la reazione di attraversamento) viene rinforzata dal lieve shock elettrico somministrato attraverso il pavimento è chiaramente un'interpretazione assai semplicistica di ciò che realmente avviene. Ciò che in effetti è rinforzato dallo shock è un certo numero di eventi, sia stimoli che risposte, da cui derivano vari condizionamenti. Poiché tutto avviene in un'unica prova, il soggetto non ha la possibilità di analizzare adeguatamente la situazione in cui viene a trovarsi. Non c'è il tempo sufficiente perché, attraverso processi di differenziazione, la proprietà di provocare determinate reazioni venga conservata solo da specifici stimoli nel caso del condizionamento classico o la puni-

Variazioni dell'attività spontanea dopo la somministrazione di uno shock elettrico dello stesso tipo di quelli usati per il condizionamento. L'attività è stata misurata contando il numero di quadrati (in cui è diviso il pavimento dell'*open field*) attraversati dall'animale in un minuto. Nella figura sono indicate le prestazioni di sette differenti gruppi di topi. L'attività dapprima è alta e poi diminuisce gradatamente in funzione dell'intervallo trascorso dopo la somministrazione dello shock. Si potrebbe così spiegare l'incubazione della reazione condizionata (*si veda la figura a pag. 69*), poiché una diminuzione dell'attività spontanea porta a un aumento della latenza della reazione di attraversamento.

zione venga associata solo a specifiche risposte nel caso del condizionamento strumentale. Così, il rinforzo costituito dallo shock si distribuisce in vario modo sui vari eventi che lo precedono.

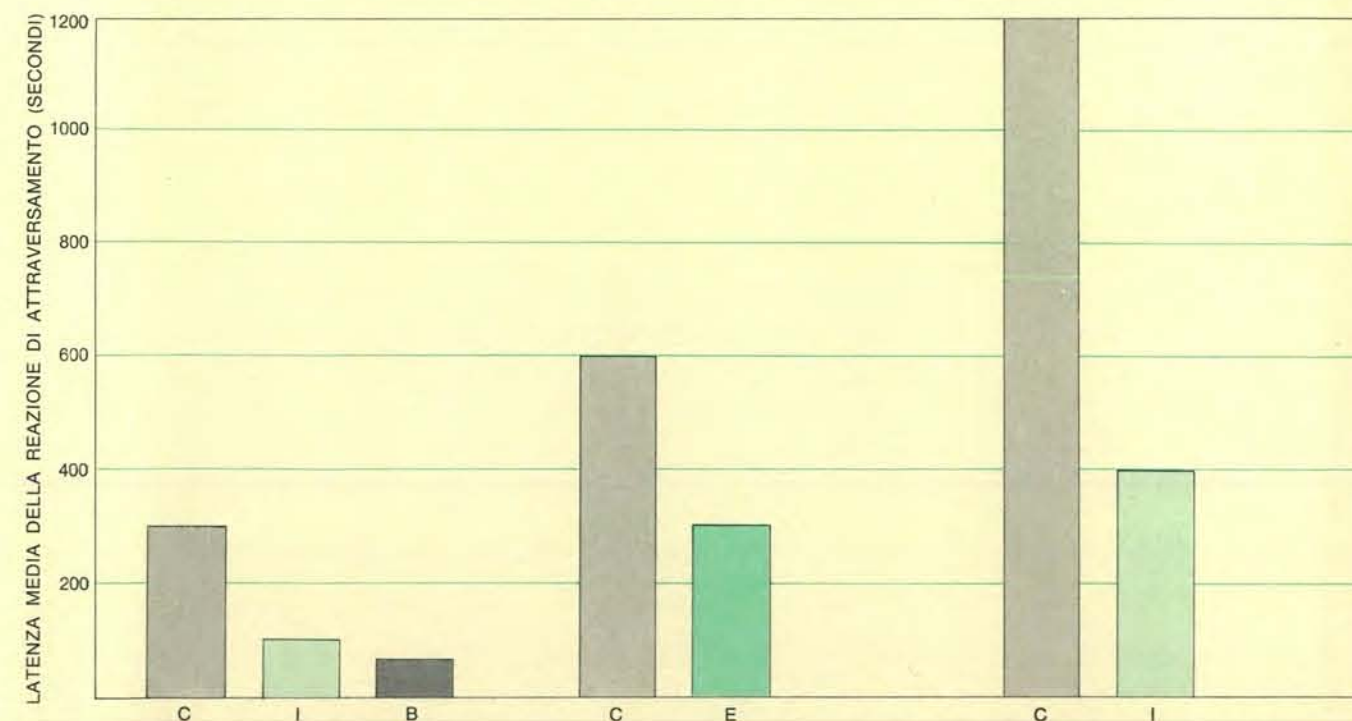
La reazione motoria di attraversamento, la cui inibizione è tradizionalmente considerata come espressione del condizionamento di evitamento di tipo passivo, è solo uno di questi eventi. In realtà, la latenza della reazione di attraversamento è, a parte gli altri fattori di cui si è già parlato, la risultante di varie reazioni condizionate, che in vari modi influiscono sul tempo che l'animale impiega, nella prova di memoria, a passare da una parte all'altra dell'apparecchio. Una reazione condizionata classica (di tipo emozionale), per esempio, si forma per il fatto che l'animale associa lo shock elettrico all'intero apparecchio, per cui è immobilizzato dalla paura quando viene rimesso nell'apparecchio stesso per la prova di memoria. Questa reazione condizionata di immobilità porta ovviamente a un allungamento della latenza della reazione di attraversamento, ma non deriva dalla specifica inibizione della reazione di attraversamento. L'animale, in questo caso, non ha paura di compiere un determinato atto motorio (quello per cui passa da una parte all'altra dell'apparecchio attraverso l'apertura di comunicazione), ma ha paura, semplicemente, e pertanto non si muove. Ciò può essere provato trattenendo l'animale in una delle due parti dell'apparecchio per un certo periodo di tempo dopo la prova di condizionamento, senza la somministrazione di shock. In questo modo la reazione condizionata emozionale all'intero apparecchio si estingue, poiché, non sopraggiungendo alcun ulteriore shock, l'animale cessa di aver paura dell'apparecchio e nella successiva prova di memoria impiega meno tempo per passare dall'altra parte (*si veda la figura a pagina 75, a sinistra*). C'è anche una generale reazione esplorativa, indipendente dagli specifici movimenti con i quali si manifesta, che viene rinforzata dallo shock. L'animale, cioè, può associare allo shock l'impulso a esplorare, e inibisce quindi questo impulso nella prova di memoria. Anche in questo caso, non ha paura di compiere un determinato atto motorio, ma ha paura di esplorare, in generale. Questa reazione condizionata può essere estinta permettendo all'animale di esplorare liberamente un ambiente nuovo (completamente diverso dall'apparecchio del condizionamento) dopo la prova di condizionamento, in modo da estinguere l'associazione fra esplorazione e shock elettrico. Di nuovo, nella successiva prova di memoria, l'animale passerà prima nella parte buia dell'apparecchio del condizionamento (*si veda la figura a pagina 75, al centro*). Inoltre, l'animale vive in gruppo normalmente e nella prova di condizionamento si trova solo, forse per la prima volta nella sua vita. L'esperienza dell'isolamento è un altro evento che può essere associato allo shock, di modo che si ha una reazione di immobilità provocata dalla paura, quando il soggetto si trova isolato. Tale reazione condizionata può essere estinta lasciando l'animale solo in una gabbia, identica a quella in cui vive, per un certo periodo di tempo dopo la prova di condizionamento, in modo da cancellare l'associazione fra isolamento e shock (*si veda la figura a pagina 75, a destra*). Insomma, l'essere trattenuti per un po' in una delle due parti dell'apparecchio, l'avere la possibilità di esplorare liberamente un nuovo ambiente, il rimanere soli, risultano essere esperienze che agiscono come eventi amnesici, che diminuiscono in vario modo la latenza della reazione di attraversamento nella prova di memoria. Né si può pensare che tale effetto amnesico sia la conseguenza dell'azione stressante di queste esperienze (la permanenza in un apparecchio in cui si sia ricevuto uno shock elettrico, l'esplorazione di un nuovo ambiente, l'isolamento, possono senz'altro essere considerate esperienze stressanti), poiché altre esperienze stressanti non hanno un effetto amnesico. Non c'è alcuna diminuzione della latenza della reazione di attraversamento, per esempio, se dopo la prova di condizionamento l'animale viene tenuto per un certo periodo di tempo in una gabbia sovraffollata.

Dunque, nel condizionamento di evitamento di tipo passivo, il passaggio nella parte buia dell'apparecchio (o gli altri tipi di risposta motoria che

vengono dati in altre versioni di questo condizionamento) non è per l'animale l'unico evento che precede la punizione e al quale quindi la punizione possa essere associata. La somministrazione dello shock elettrico è preceduta da altri eventi e il suo effetto « rinforzante » si distribuisce in un modo o nell'altro su di essi, forse concentrandosi su quegli eventi a cui il soggetto ha prestato maggiore attenzione. Così per un soggetto potrebbe essere determinante l'esperienza della reazione motoria, per un altro l'esperienza dell'esplorazione, per un altro ancora l'esperienza dell'isolamento, per un altro infine l'esperienza di quel particolare ambiente (cioè l'apparecchio).

È facile capire quanto tutto ciò renda difficile lo studio dei processi della memoria. Se, dunque, non è possibile parlare di una reazione condizionata, ma è necessario ammettere la formazione simultanea di più reazioni condizionate, senza essere in grado oltretutto di stabilire se esse siano sempre presenti in ogni soggetto, lo studio dei processi della memoria dovrebbe presupporre: 1) un criterio per stabilire quali reazioni condizionate si siano formate; 2) la conoscenza degli andamenti temporali di ciascuna reazione; 3) la conoscenza della vulnerabilità di ciascuna reazione ai vari trattamenti. Per ora non è possibile avere nessuna di queste informazioni. È anche opportuno ricordare che l'effetto amnesico che si verifica quando i soggetti vengono tenuti per un certo periodo di tempo nell'apparecchio del condizionamento è stato ottenuto anche nel condizionamento di evitamento di tipo attivo, che richiede molte prove. Probabilmente la spiegazione migliore consiste nel fatto che la caratteristica del condizionamento che si realizza in un'unica prova, e cioè l'impossibilità in cui si trova il soggetto di percepire chiaramente la situazione e quindi di associare il rinforzo a specifici aspetti di tale situazione, può riscontrarsi anche nella fase iniziale di un apprendimento che si realizzi in molte prove. In questa fase iniziale, quindi, si possono formare molte reazioni condizionate « occasionali » (si pensi al comportamento « superstizioso » di cui parla Skinner), che saranno poi estinte attraverso i processi di differenziazione. Ma si deve anche prendere in considerazione la possibilità che alcune di queste reazioni condizionate occasionali rimangano come componenti permanenti della cosiddetta reazione condizionata di evitamento di tipo attivo.

Quanto è stato detto si può riassumere in tre punti fondamentali. Primo: è necessario che apprendimento e memoria vengano sempre studiati insieme, evitando ogni artificiale distinzione delle due aree di ricerca. Secondo: nuovi modelli della memoria devono essere elaborati, i quali prendano in considerazione tutte le fasi del processo della memoria, dalla registrazione alla codificazione, dal consolidamento alla ritenzione, al recupero, e a qualsiasi altra fase la cui esistenza possa essere suggerita dai dati sia sperimentali che osservazionali. Terzo: i comportamenti devono essere sempre analizzati considerando tutti i fattori che possono averli determinati e, quindi, in tanto si può assumere che un certo comportamento misuri l'apprendimento o la memoria in quanto si sia capaci di valutare il contributo che a quel comportamento hanno eventualmente dato altri fattori, diversi dall'apprendimento e dalla memoria.

C'è, però, collegato a quest'ultimo, un altro punto importante su cui vale

Trattenendo i soggetti per 10 minuti nella parte illuminata (*I*) o nella parte buia (*B*) dell'apparecchio subito dopo la prova di condizionamento, si ha una parziale estinzione della reazione condizionata di evitamento di tipo passivo. I soggetti del gruppo di controllo (*C*) sono stati tolti dall'apparecchio subito dopo la prova di condizionamento. Nel grafico, a sinistra, sono indicate le prestazioni degli animali (topi) nella prova di memoria, effettuata 24 ore dopo la prova di condizionamento. Si ha una parziale estinzione della reazione condizionata di evitamento di tipo passivo permettendo agli animali di esplorare liberamente per 10 minuti un ambiente nuovo subito dopo la prova di condizionamento (*E*). I soggetti del gruppo di controllo (*C*) sono stati rimessi, subito dopo la prova di condizionamento, nella gabbia in cui vivono normalmente. Nel grafico, al centro, sono indicate le prestazioni degli animali (topi) nella prova di memoria, effettuata 24 ore dopo la prova di condizionamento. Si ha una parziale estinzione della reazione condizionata di evitamento di tipo passivo lasciando isolati gli animali in una gabbia identica a quella in cui normalmente vivono in gruppo, per 10 minuti dopo la prova di condizionamento (*I*). I soggetti del gruppo di controllo (*C*) sono stati rimessi, subito dopo la prova di condizionamento, nella gabbia in cui vivono. Nel grafico a destra, sono indicate le prestazioni dei topi nella prova di memoria, effettuata 24 ore dopo la prova di condizionamento.

la pena di spendere qualche parola. Si è verificata negli ultimi anni una certa tendenza, da parte di alcuni studiosi del comportamento animale, a dilatare arbitrariamente le aree di ricerca della memoria e dell'apprendimento fino a portarle a fondersi in qualche modo con quella dell'intelligenza, operazione estremamente pericolosa per le sue implicazioni al livello di un'eventuale estrapolazione alle corrispondenti aree della psicologia umana. È bene chiarire che la psicologia animale non ha per ora né i modelli teorici né le metodologie per stabilire che cosa si possa intendere per « generale capacità di apprendimento » in un animale. Appare, quindi, assolutamente problematica per ora la definizione di « intelligenza » di un animale e del tutto impensabile qualsiasi tentativo di riferimento all'intelligenza umana. Da un lato, quasi tutti gli studi effettuati dimostrano una mancanza di correlazione fra differenti test di apprendimento. Cioè, i soggetti che riescono meglio in un test non sono necessariamente quelli che riescono meglio in un altro test. Non esistono, dunque, da questo punto di vista, individui che possano essere definiti « piú intelligenti » di altri. Tutto dipende dal tipo di apprendimento. Ciò non deve stupire, ovviamente, dato che, come è stato detto piú volte, le prestazioni nei test di apprendimento misurano anche altri fattori oltre alla capacità di apprendimento. D'altro lato, non bisogna dimenticare che qualsiasi eventuale tentativo di definire l'intelligenza di un animale deve basarsi sulla complessità dei problemi risolti e non sulla rapidità di soluzione di certi problemi. Questo genere di ricerca, però, risulta purtroppo difficilissimo, fino al punto che la psicologia comparata l'ha ormai quasi completamente abbandonato. Date le differenze al livello dei meccanismi sensori, motori, delle motivazioni, etc., è pressoché impossibile creare delle situazioni che ci permettano di valutare in senso differenziale la capacità di apprendimento di due specie animali. Ma questa difficoltà esiste anche quando si vogliano paragonare due individui della stessa specie. Per valutare le loro rispettive intelligenze dovremmo al solito essere in grado di scomporre ogni loro prestazione nelle varie componenti riferibili ai vari fattori, innati o acquisiti, che determinano quella prestazione. In questa luce vanno visti, ad esempio, molti risultati di psicogenetica animale. Non si può parlare di eredità di certe « capacità di apprendimento » nei casi in cui si sia semplicemente messa in evidenza una eredità della « capacità di fornire un certo tipo di prestazione in una data situazione ». Infatti, ciò che è in realtà ereditato potrebbe essere un determinato carattere psicologico o fisiologico, diverso dalla capacità di apprendimento, il quale risulti determinante nella prestazione esaminata. In questo caso il riconoscimento della complessità dei fattori che determinano una prestazione e la non identificazione di questa, quindi, con il risultato della capacità di apprendimento, è molto importante, perché tale identificazione porta facilmente alla tesi dell'ereditarietà dell'intelligenza, che pertanto deve considerarsi una conclusione sbagliata che deriva da una serie di premesse sbagliate. In una prospettiva meno sensazionalistica, ma certamente piú scientifica, lo studio dei processi dell'apprendimento e della memoria negli animali deve servire solo, per ora, per analizzare certi aspetti fondamentali di questi processi, cioè certi meccanismi del sistema nervoso che, conservando gli effetti di un'esperienza, portano alla modificazione del comportamento.

Sarebbe inoltre opportuno che gli studiosi si rendessero conto che fenomeni complessi come l'apprendimento e la memoria non possono essere studiati in un'unica prospettiva. In questo articolo si è parlato solo della ricerca psicologica, ma questo non implica certo la convinzione da parte dell'autore che questo tipo di ricerca esaurisca tutto il campo sterminato degli studi sull'apprendimento e sulla memoria, né che esso rappresenti il modo migliore di portare avanti questi studi. L'estrema complessità dei fenomeni dell'apprendimento e della memoria esige una ricerca interdisciplinare, in senso sia esterno che interno alla psicologia. Da un lato, cioè, la psicologia deve allearsi ad altre scienze, soprattutto alla fisiologia, alla farmacologia, alla biochimica, all'etologia. Dall'altro, è necessario che nell'ambito stesso della psicologia si realizzi un'integrazione dei vari approcci e delle varie tendenze. I metodi cosiddetti obbiettivi sviluppati dal pavlovismo o dal behaviorismo devono essere usati parallelamente a quelli della psicologia cognitivistica e di quella dinamica, e inoltre deve essere realizzato un maggior dialogo fra area sperimentale e area clinica. Di fondamentale importanza, infine, deve essere lo sforzo per tradurre in un linguaggio comune le conclusioni teoriche delle diverse scienze o delle diverse tendenze. Oggi il linguaggio è cosí diverso nei vari casi che un behaviorista e un pavloviano, ad esempio, hanno poche probabilità di capirsi, anche quando parlano della stessa cosa.

La fionda come arma

Questo strumento manuale non è noto quanto l'arco, malgrado venisse usato dalle truppe equipaggiate con armi leggere, sia in India e in Persia che in Grecia e a Roma, e sopravvivesse persino all'avvento della polvere da sparo

di Manfred Korfmann

Tutti sanno che Davide uccise Golia con una fionda, ma qual è il posto occupato dalla fionda nella storia della tecnologia? Il fatto è che le fionde venivano impiegate come armi belliche regolari in Europa e nel Vicino Oriente a partire almeno dall'Età del Bronzo fino al XVII secolo della nostra era. Inoltre la fionda è stata, tra le armi a lunga gittata, la favorita per molti popoli, del passato e del presente, in tutto il mondo. In Mesopotamia, in Persia, in Grecia e a Roma un fromboliere non era considerato da meno di un arciere. In questa parte del mondo la fionda era nota probabilmente fin dall'inizio del Neolitico, circa 10 000 anni fa, e potrebbe essere stata usata anche verso la fine del Paleolitico.

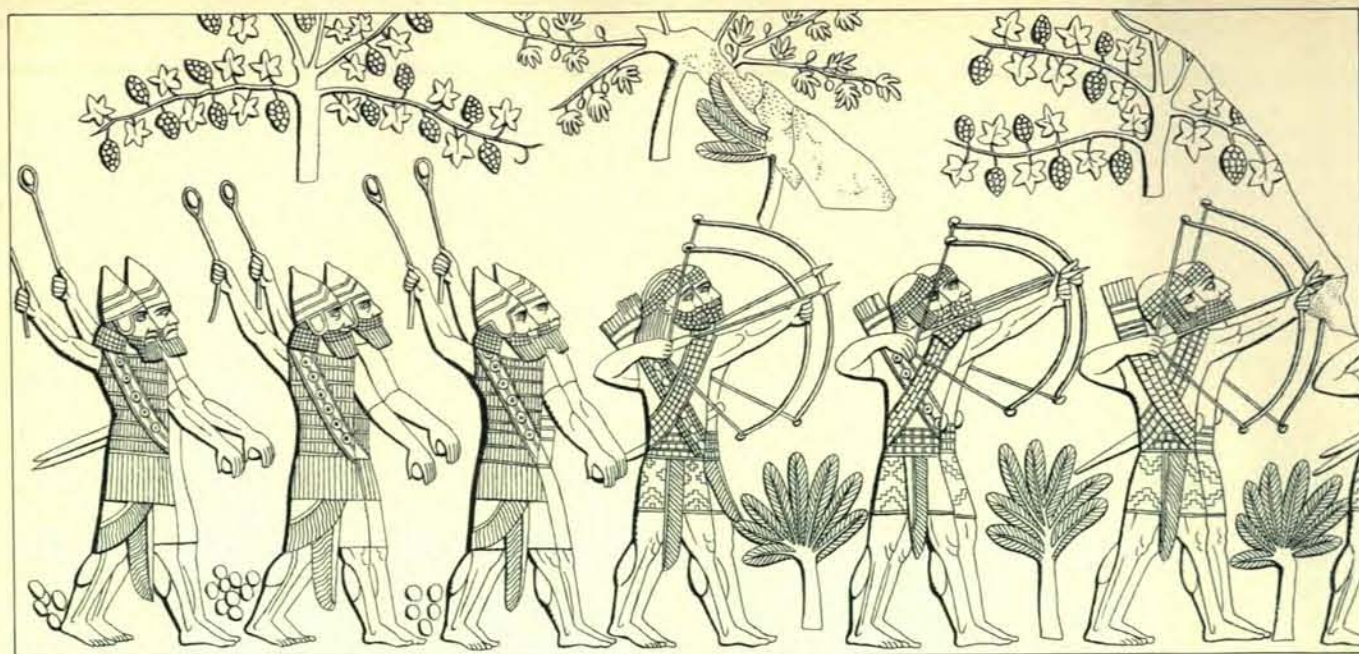
La vittoria di Davide su Golia è sempre stata vista sotto il profilo allegorico, ma pensando a come si svolgeva la guerra a quei tempi andrebbe piuttosto considerata come un esempio sia della grande destrezza posseduta dai frombolieri sia della fiducia che essi riponevano nella loro arma. Un esame di quel famoso scontro, come è riportato nel primo libro di Samuele, avvalora tale punto di vista. Davide, si ricorderà, era l'ottavo figlio di Iesse. Essendo il più giovane, badava alle pecore, e tale occupazione conferma la sua familiarità con la fionda: infatti questa era l'arma impiegata a quell'epoca dal pastore per proteggere i propri animali. Davide era poi stato destinato a suonare l'arpa e a portare le armi di Saul, il re d'Israele ai tempi in cui gli ebrei erano in guerra con i filistei. Ogni giorno un campione filisteo, Golia di Gat « che era alto sei cubiti e una spanna », usciva dal campo nemico e proponeva un combattimento con qualsiasi campione israelita di modo che il vincitore del duello decidesse le sorti della guerra.

Davide si recò all'accampamento israelita dopo che per quaranta giorni di seguito Golia aveva lanciato la sua sfida senza che questa venisse raccolta da un solo guerriero ebreo. Davide si offrí di combattere contro Golia, ma rifiutò di indossare l'armatura o di portare le armi che Saul voleva imporgli. Cosí prosegue a questo punto la narrazione: « Egli prese in mano il suo bastone (di pastore) e si scelse cinque pietre lisce dalla valle del torrente e le mise nella sua borsa da pastore che gli serviva da recipiente... e in mano aveva la sua fionda: e si accostò al filisteo. Quindi Davide mise la mano nella sua borsa e vi prese una pietra e la frombolò cosí che colpí il filisteo sulla fronte e la pietra gli affondò nella fronte ed egli cadde con la faccia a terra. E Davide continuò a correre e stette sul filisteo, quindi gli prese la spada e gliela trasse dal fodero e definitivamente lo mise a morte quando con essa gli troncò la testa. E i filistei videro che il loro potente era morto e si diedero alla fuga ».

A Golia era stata fornita per lo scontro un'armatura completa: un elmo di metallo, una cotta di maglia di ferro, gambali metallici ai piedi e un piccolo scudo appeso al braccio. Un portatore lo precedeva reggendo uno scudo di dimensioni piú grandi. La spada che Davide usò per tagliare la testa di Golia non è descritta, ma la lancia del filisteo aveva un'asta « simile al subbio dei tessitori » e una pesante lama. Questo equipaggiamento non è dissimile da quello adottato dall'oplita greco o dalla fanteria pesante. Esso è destinato a una strategia d'urto: la lancia non deve essere scagliata, ma impugnata per trafiggere o come difesa dall'attacco della cavalleria. L'armatura e le armi di Golia, se si eccettua forse la lancia, erano anche appropriate per un duello con un altro guerriero equipaggiato allo stesso modo, ma del tutto inadatte per l'inseguimento di un avversario disarmato e dal piede veloce come Davide, a cui bastava mantenere le distanze per essere completamente al sicuro.

Davide non aveva alcuna intenzione di portarsi a stretto contatto col proprio avversario, poiché la fionda è un'arma a lunga gittata. Comunque, per quanta fiducia egli potesse riporre nell'aiuto di Dio, nel prepararsi per lo scontro non scelse un solo sasso, bensí cinque. Se con il suo primo tiro non avesse colpito Golia in viso – il punto vitale non protetto a cui Davide aveva sicuramente mirato – avrebbe avuto a sua disposizione altri quattro sassi. Tutto sommato sembrerebbe giusto attribuire la vittoria di Davide piú alla sua abilità come fromboliere che all'intervento divino.

Davide è il fromboliere biblico piú noto, ma non è il solo. I frombolieri mancini di Beniamino (*Giudici*) provocarono gravi perdite tra gli israeliti, e gli stessi corpi scelti di Davide « potevano usare sia la mano destra sia la mano sinistra nel "lanciare" le pietre » (*Cronache*). Perché allora la fionda è passata quasi inosservata come arma bellica? Qualcosa che somiglia a una risposta si trova nell'*Iliade*. Nel racconto di Omero dell'assedio di Troia viene fatto cenno ai locresi, guerrieri senza armatura che confidavano nel

Nella pagina a fronte: Davide e Golia come appaiono effigiati sulla parete di una chiesa armena del X secolo su un'isola del lago Van. Golia è rappresentato con la spada sguainata. Davide, con la sua fionda carica e pronta, è in piedi, in questo ritratto, troppo vicino a Golia; un lancio di 225 metri non era insolito per i frombolieri di quell'epoca.

In questo disegno, basato su un rilievo di Ninive raffigurante una delle campagne di Sennacherib (704-681 a.C.), i frombolieri assiri con le loro fionde si trovano dietro gli arcieri. La loro posizione fa ritenere che superassero in gittata gli arcieri.

Un ausiliario romano delle guerre daciche, la fionda in posizione di tiro, tiene i proiettili di scorta nella piega del mantello buttato sul braccio sinistro. Le figure di questo disegno appartengono alla colonna eretta per commemorare le vittorie di Traiano.

loro arco e nella « treccia di lana di pecora ». La parola greca che sta per « fionda » tuttavia compare una sola volta in tutto il poema, e comunque vi vien fatto cenno non come a un'arma, bensí come a una benda improvvisata: un guerriero troiano fascia la mano ferita di un compagno « con buona treccia di lana di pecora, con la fionda, che lo scudiero portava al pastore di popoli ». Sembrerebbe che i peltasti greci, le truppe dotate di armi leggere tra le cui fila militavano i frombolieri, i lanciatori di giavellotto e gli arcieri, godessero di scarsa considerazione in tempi in cui era ritenuto onorevole solo il combattimento faccia a faccia tra guerrieri rivestiti di armatura. Nell'arte figurativa le rappresentazioni di opliti (e persino di arcieri e lanciatori di giavellotto) sono abbastanza comuni, ma quelle dei frombolieri sono rare.

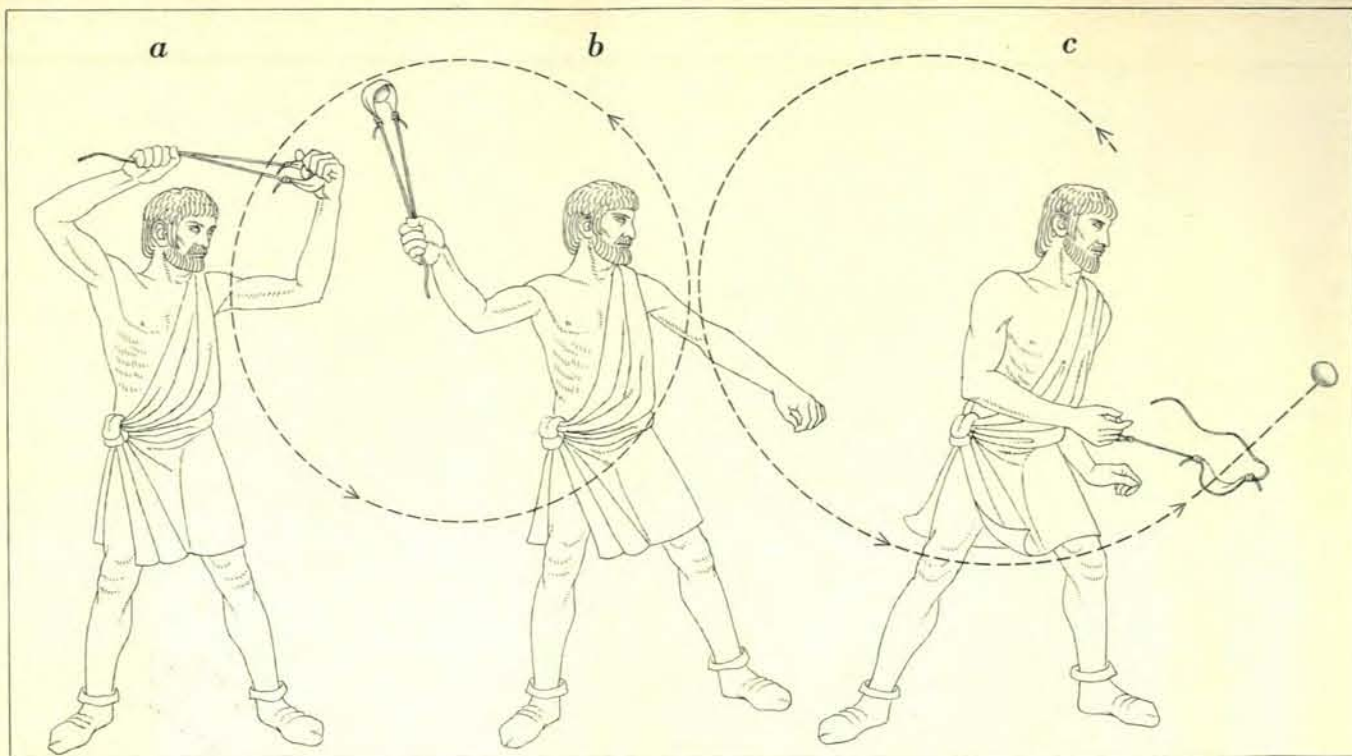
Nel periodo greco classico tuttavia le truppe leggere avevano un ruolo di primo piano nel combattimento. Loro era il compito di dar inizio alla battaglia. Il diluvio di giavellotti, frecce, sassi da essi provocato doveva aprire una breccia nei ranghi nemici; come minimo avrebbe messo in luce i punti deboli della formazione avversaria che sarebbero poi stati sfruttati dalla fanteria pesante. Inoltre, se l'assalto non avesse avuto successo, le truppe leggere sarebbero state in grado di coprire la ritirata della fanteria pesante. Un'armata che ingaggiasse battaglia senza i peltasti poteva già considerarsi sconfitta.

Possediamo un resoconto particolareggiato di quanto successe a un esercito di tal fatta, che era rimasto privo di quasi tutte le truppe leggere. Si trattava del reparto di 10 000 fanti greci dotati di armi pesanti, l'ossatura di un esercito molto piú grande che aveva tentato di rovesciare il re di Persia nel 401 a.C. Dopo che il pretendente al trono che li aveva guidati era caduto nella battaglia di Cunassa, le sue truppe indigene si erano date alla fuga e i greci si erano trovati soli. L'ateniese Senofonte si incaricò di guidare verso la salvezza i 10 000 fanti greci, ma fin dal primo giorno di marcia essi furono cosí infastiditi da un piccolo numero di cavalieri, arcieri e frombolieri nemici che poterono percorrere solo circa 25 stadi: poco piú di quattro chilometri e mezzo. Quella notte Senofonte dichiarò ai suoi capitani: « Abbiamo bisogno immediatamente anche noi di frombolieri e di una cavalleria ».

« Ora io dico – continuò Senofonte – che nel nostro esercito ci sono dei rodi, la maggior parte dei quali sa usare la fionda e riesce a lanciare i propri proiettili a una distanza perlomeno doppia rispetto a quella coperta dalle fionde persiane. Infatti queste ultime hanno una gittata corta poiché le pietre che esse lanciano sono grandi soltanto quanto possono essere contenute in una mano; i rodi invece sono abili anche nell'arte di lanciare proiettili di piombo.

Posizione di partenza, con la fionda in equilibrio sopra la testa, in una scultura raffigurante un fromboliere neoittita di Tell Halaf in Siria, eseguita nel IX o nell'VIII sec. a.C.

Senofonte e i suoi colleghi scelsero tra i ranghi 200 frombolieri e una squadra di 50 cavalieri che montavano cavalli da soma. L'aggiunta di queste forze dotate di armi leggere a un gruppo di 200 arcieri cretesi scelti tra i 10 000, permise da quel momento in avanti ai greci di farsi onore nel rintuzzare i persiani. La gittata degli arcieri cretesi non era paragonabile a quella dei persiani, ma i frombolieri di Rodi, sottolineava Senofonte, « lanciavano i loro proiettili piú lontano dei persiani, addirittura piú lontano di quanto non riuscissero a fare gli arcieri nemici ». Se si considera che gli arcieri persiani erano ritenuti a quei tempi i migliori del mondo, tale giudizio attesta chiaramente l'abilità dei frombolieri greci.

Ma, con esattezza, come si possono confrontare, quanto a gittata, la fionda e l'arco? Il cronista militare romano Vegezio, che scriveva all'incirca nel 400 d.C., consigliava di esercitarsi con l'arco a una distanza di 180 metri. Anche un moderno arco da competizione, con una trazione di 20 chilogrammi, può scagliare una freccia a una distanza di poco superiore ai 200 metri. Con

Il lancio di un fromboliere comincia (*a*) con il caricamento e il bilanciamento della fionda; il fromboliere infila un dito nel cappio dell'estremità fissa della fionda tenendo l'estremità libera tra il pollice e l'indice. Tre o quattro rotazioni della fionda in senso antiorario (*b*), effettuate essenzialmente con un movimento del polso piuttosto che del braccio, imprimono al proiettile la massima velocità. Il proiettile viene scagliato (*c*) quando il fromboliere abbandona l'estremità libera della fionda; all'inizio della sua traiettoria parabolica il proiettile ha una velocità iniziale di circa 100 chilometri orari.

una freccia lunga e leggera e una trazione di 26 chilogrammi, un arciere che miri solo alla distanza può raggiungere all'incirca i 275 metri. Per avere un termine di paragone, chiesi ad alcuni giovani della Turchia orientale di lanciare per me, con la fionda, dei comunissimi sassi. In cinque prove su undici, i sassi raggiunsero il suolo al di là una linea di demarcazione posta a una distanza di 200 metri: i tre lanci migliori si collocarono tra i 230 e i 240 metri. Nessuno di quei giovani risultava essere un tiratore di fionda specializzato. Oltre a tutto, come proietli erano stati impiegati dei ciotoli scelti a caso invece che le pietre accuratamente forgiate o i proiettili di terracotta e di piombo lanciati dai frombolieri dell'epoca greca e romana. Se ci si basa solo sul commento di Senofonte, sembra probabile che un fromboliere dotato di proiettili di piombo potesse scagliarli fino a una distanza, valutata per eccesso, di 400 metri.

La fionda di cui stiamo parlando è la normale fionda fatta di materiale elastico (*funda* in latino), ma ve n'è anche un altro tipo detto fustibalo (*fustibalus* in latino). La prima può anche essere costituita solo da una cinghia lunga approssimativamente 90 centimetri e larga circa 2,5 centimetri. A un'estremità la cinghia presenta un cappio o un nodo o un fiocco che consentono di fissarla a un dito qualsiasi della mano che il fromboliere usa per il lancio. L'altra estremità, che può essere annodata per fornire una presa, viene afferrata tra il pollice e l'indice della stessa mano. Il lanciatore colloca il proiettile in una specie di « tasca », talvolta appositamente allargata, alla estremità del cappio penzolante. Se il proiettile è fatto di pietra o di terracotta ha di solito la forma di un piccolo uovo. Il movimento rotatorio del polso può far roteare rapidamente la fionda sia orizzontalmente (intorno alla testa del fromboliere) sia verticalmente (in un piano parallelo al corpo). Dopo averle fatto compiere tre o quattro giri, il fromboliere abbandona l'estremità libera della cinghia e il proiettile viene scagliato via secondo la tangente al cerchio percorso dalla fionda.

Il fustibalo ha una gittata inferiore alla fionda, d'altra parte è piú facile da maneggiare e può essere impiegato per lanciare proiettili piú grandi e pesanti. La fionda, fatta solitamente di corda, viene legata a un'estremità a un'asta che può essere lunga anche 90 centimetri. L'estremità libera viene anch'essa temporaneamente collegata alla punta dell'asta; la parte terminale dell'asta può presentare una tacca in modo che l'estremità della fionda possa liberarsi, oppure a quest'ultima viene fatto un cappio che al momento opportuno possa sfilarsi agevolmente. Una tasca profonda contiene il proiettile. L'asta, che in un primo tempo è tenuta parallela al terreno, viene portata bruscamente in posizione verticale sopra la testa del fromboliere; quando il movimento rotatorio ha raggiunto la massima intensità, l'estremità libera della fionda si sfila e il proiettile viene scagliato lontano (*si veda l'illustrazione della pagina a fronte*). Impiegato sia in epoca greca sia in epoca romana, il fustibalo fu un'arma molto popolare durante gli assedi per tutto il Medioevo. Anche con l'avvento della polvere da sparo fu utilizzato per lanciare le granate per buona parte del XVII secolo.

Quanto piú lunga è la fionda, tenendo naturalmente conto dei limiti pratici, tanto piú grande è la velocità potenziale del proiettile. Gli indigeni delle isole Baleari, che si trovano a sud-est della Spagna, erano frombolieri abilissimi. Infatti Polibio, uno storico greco del II secolo a.C., affermava che tali isole dovevano il loro nome proprio a questo fatto, giacché *ballein* era la parola greca che significava « lanciare ». Comunque sia, nei tempi classici i frombolieri balearici vennero impiegati come truppe leggere in molte guerre, la piú nota delle quali è forse il lungo conflitto tra Roma e Cartagine. Essi portavano sempre con sé fionde di tre lunghezze diverse a seconda della gittata che desideravano ottenere: una lunga, una corta e una intermedia.

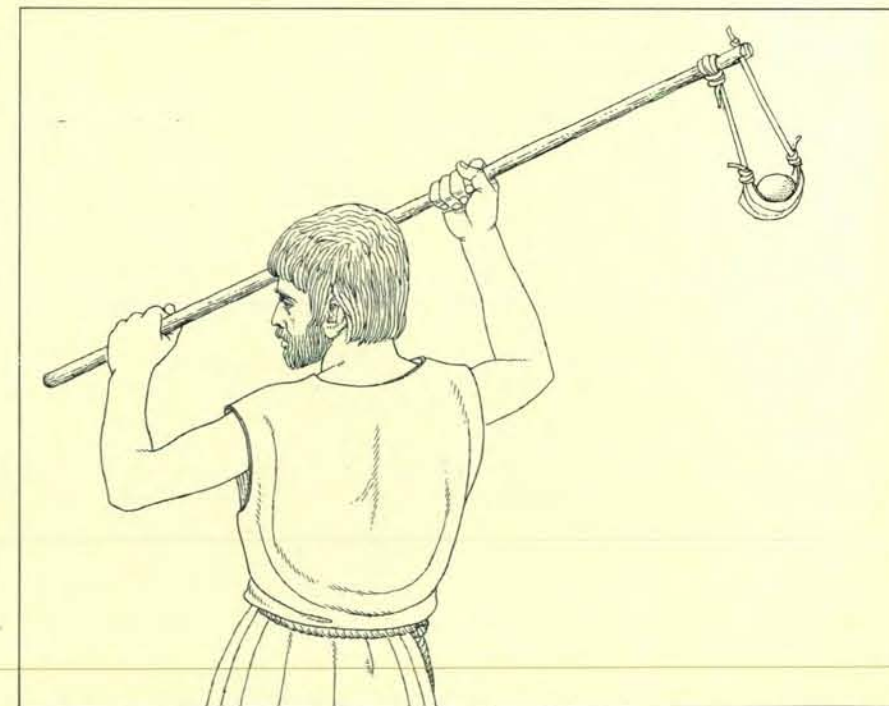
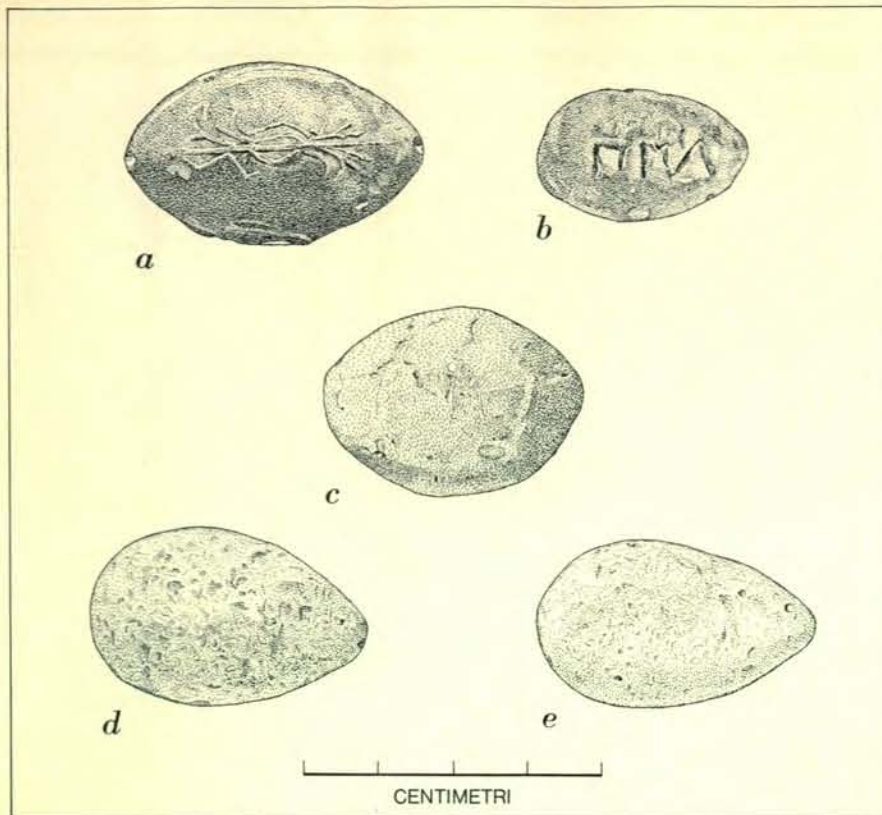
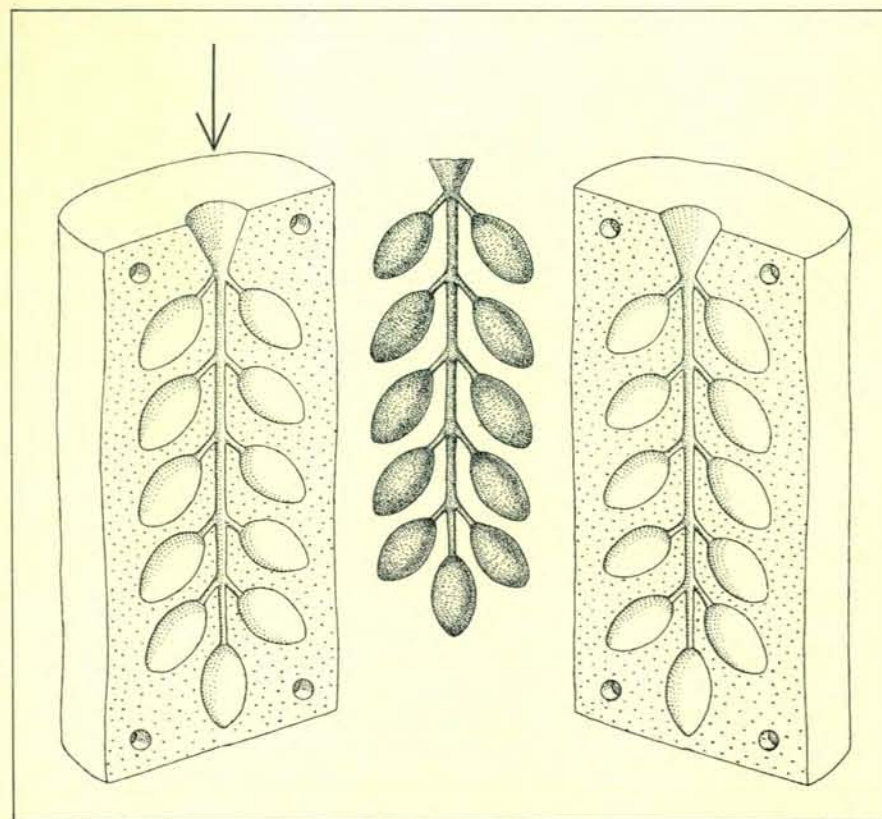
Passando ai proiettili, è chiaro che quando non si tratta che di semplici ciottoli levigati dall'acqua difficilmente gli archeologi riescono a identificarli come tali. Solo quando un gran numero di pietre simili vengono trovate insieme e non vi è alcuna prova che fossero destinate ad altro scopo (ad esempio per martellare o levigare qualcosa) o, meglio ancora, quando sono di una qualità estranea al posto in cui sono state rinvenute, possono essere considerate un indizio potenziale dell'uso di fionde in quel luogo e in quel tempo. Fortunatamente, sebbene sia impossibile riconoscere molti proiettili in quanto forgiati solo dalla natura, un buon numero di essi, nei tempi antichi, veniva fabbricato con gran cura. Non è sempre facile riconoscere tali manufatti: anche quando ciò è avvenuto, spesso gli archeologi si sono posti l'interrogativo di quale fosse lo scopo di tali strane « uova di terracotta ».

Nel Vicino Oriente, i primi proiettili da fionda fatti a mano erano sferici e avevano fatto la loro comparsa poco prima dell'inizio del VI millennio a.C. Subito dopo apparvero i proiettili biconici (*si veda l'illustrazione a pagina 84 in alto*). Un millennio dopo, all'incirca nel 4000 a.C., venivano impiegati anche proiettili di forma ovoidale. Al processo evolutivo e alla standardizzazione dei tipi di proiettili sembra abbiano concorso tre considerazioni, tutte relative alla ricerca di una sempre maggior perfezione. Il primo obiettivo era quello di rendere il peso dei proiettili ragionevolmente uniforme, di modo che il fromboliere non dovesse, a ogni lancio, compensarne la diversità. Il secondo obiettivo consisteva nel dar loro una forma costante, in certo modo aerodinamica, nell'interesse non solo della precisione, ma anche della velocità e della distanza. Il terzo obiettivo mirava ad avere il proiettile adatto alla tasca della fionda di modo che, come sottolineava lo storico Livio, « il proiettile non volasse a caso... ma, trattenuto saldamente mentre veniva fatto roteare, potesse venir lanciato come da un arco ».

Nel fabbricare questi proiettili di pietra dalla forma voluta, gli artigiani mostrarono una naturale preferenza per quei materiali facilmente lavorabili come il calcare. Tuttavia, in un'epoca antichissima, ancora piú antica della fase preceramica del Neolitico, alcuni uomini si resero conto che, come materiale, la ceramica presentava dei vantaggi. In tutto il mondo, in siti sia preistorici sia storici, sono stati rinvenuti proiettili di terracotta: ad esempio ne sono stati trovati alcuni, vecchi di 7000 anni, ad Hassuna, in Iraq, mentre altrove ne sono stati portati alla luce non a centinaia ma a migliaia. Non fu la mancanza della pietra adatta a favorire l'uso dell'argilla: si sono trovati infatti proiettili di terracotta anche in molti luoghi dove l'approvvigionamento di ciottoli era estremamente agevole.

I proiettili di terracotta sono curiosi per due aspetti. In primo luogo, in quasi tutti i casi essi sono stati fatti indurire mediante essiccazione al sole piuttosto che in forni. In secondo luogo, sono sorprendentemente pesanti in rapporto alle loro dimensioni. I due fatti sono in stretto rapporto fra loro. Al fine di ottenere il massimo del peso all'interno di certi limiti dimensionali, coloro che fabbricavano questi proiettili non mescolavano all'argilla la paglia come facevano invece per vasi e mattoni: i proiettili erano costituiti di argilla pura (o, in rarissimi casi, di ciottoli rivestiti di argilla) e avevano una densità corrispondente. Se questi proiettili di argilla pura fossero stati cotti al fuoco, il calore li avrebbe spezzati e resi inutilizzabili: ecco perché venivano fatti asciugare lentamente al sole.

Ai tempi della Grecia classica, se non prima, era divenuto d'uso comune un altro tipo di proiettili, quelli fatti di piombo. I romani li chiamavano *glandes* a causa della loro somiglianza con le ghiande: venivano fusi in stampi e spesso recavano delle iscrizioni. Le stesse poche lettere incise nella parte cava dello stampo potevano lasciare la loro impronta su molte centinaia di proiettili. Spesso le iscrizioni erano di ordinaria amministrazione: il nome o il numero della formazione militare del fromboliere, il nome dello stato in guerra o quello del condottiero. Alcuni tuttavia erano meno formali. Uno porta scritto « prendi questo », un altro si vanta « un soffio acheo », su un terzo si legge « il tuo cuore a Cerbero », su un quarto « per il didietro di Pompeo » e su un quinto, semplicemente, « ahimé ».

Come suggerisce il confronto fatto da Senofonte tra i proiettili rodi e quelli persiani, i proiettili da fionda erano molto diversi per dimensioni e peso. Misurando un campione rappresentativo costituito da proiettili di pietra biconici e ovoidali, provenienti da località del Vicino Oriente, è possibile rendersi conto di quanto ampia fosse la lo-

Il lancio effettuato con il fustibalo consiste in un brusco passaggio da una posizione orizzontale a una verticale. L'estremità libera della corta fionda viene infilata in una tacca praticata sulla punta dell'asta e si scioglie quando il movimento rotatorio ha raggiunto il massimo scagliando il proiettile. Il fustibalo ha una gittata minore della fionda, ma permette di lanciare proiettili piú pesanti e spesso era usata per le granate.

Tra questi proiettili artificiali, ben diversi dai ciottoli, ve ne sono due di piombo (*a* e *b*). Quello piú grande, di origine greca o romana, presenta un motivo a saetta, quello piú piccolo è stato trovato negli scavi di Olinto in Grecia. Il terzo proiettile (*c*) è di tipo biconico ed è fatto di argilla essiccata. Gli altri (*d*, *e*) sono ovoidi di pietra.

Stampo di terracotta del tipo « ad albero » usato per fondere undici proiettili di piombo. Questa ricostruzione è basata su un frammento di stampo rinvenuto a Olinto. È raffigurata la forma ad albero che si ottiene prima che vengano staccati i proiettili.

ro gamma. Il peso minimo era di 13 grammi, quello massimo di 185 grammi. Quanto al volume, si andava dai 5 ai 65 centimetri cubi (se i proiettili fossero stati perfettamente sferici, i diametri corrispondenti a questi volumi sarebbero stati rispettivamente di due e cinque centimetri).

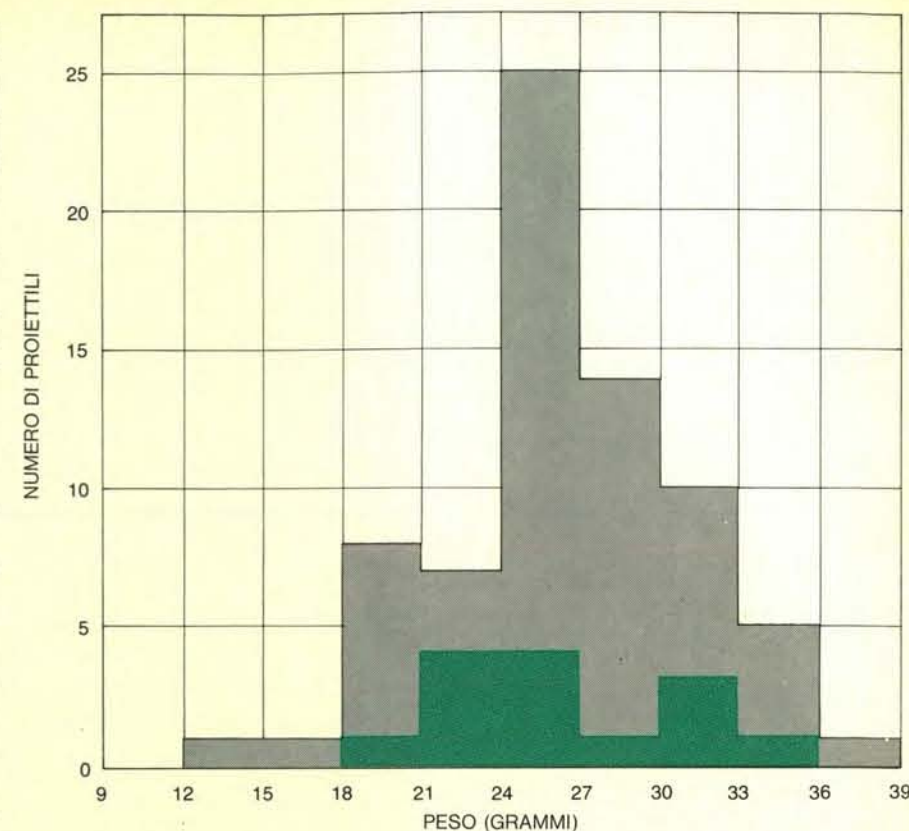
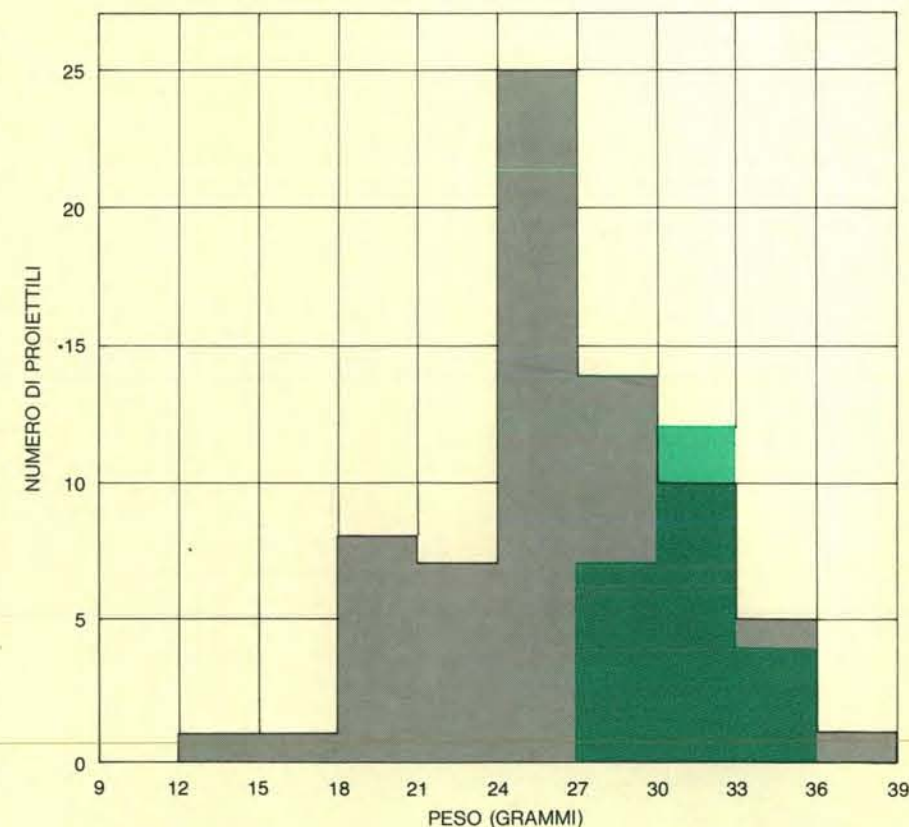
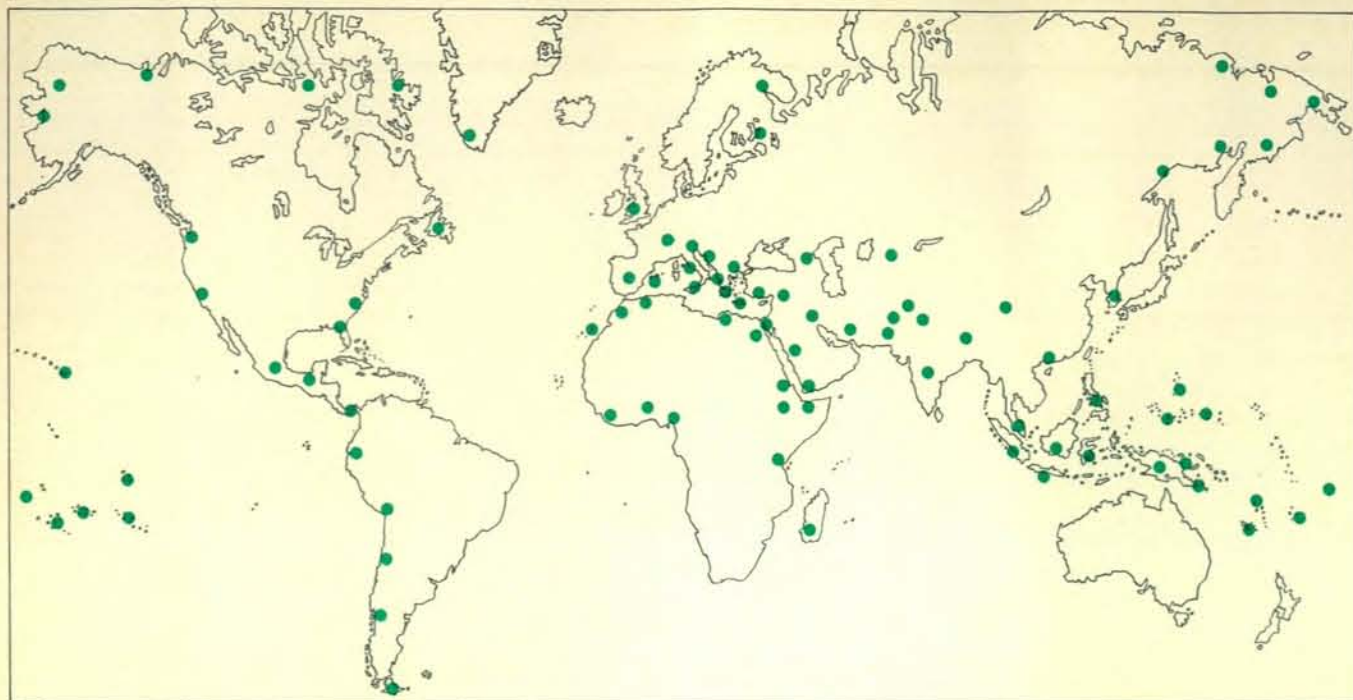
Considerando i proiettili da fionda in generale, fossero essi fatti di pietra, di terracotta o di piombo, il loro peso è generalmente intermedio a quello appena citato. Pochi dei proiettili piú leggeri pesano meno di 20 grammi e pochi dei piú pesanti superano i 50 grammi. Questo era vero, per esempio, in epoca romana. Nel 1885 lo studioso di antichità classiche K. Zangemeister pubblicò, fra l'altro, dei dati relativi al peso di proiettili provenienti da alcune località della Sicilia e dell'Italia. Egli aveva trovato che i proiettili dal peso mediamente inferiore (che andavano da un minimo di 24 grammi a un massimo di 46 grammi) provenivano dalla Sicilia. I piú pesanti provenivano da Ascoli, sulla penisola, e il loro peso medio superava i 47 grammi. Quelli che provenivano da una seconda località della penisola, Perugia, avevano un peso intermedio.

I proiettili usati talvolta dai frombolieri balearici costituiscono un'importante eccezione persino del massimo di 185 grammi riscontrato nel Vicino Oriente. Lo storico di origine siciliana Diodoro Siculo, che scriveva nel I secolo a.C., diede un preciso resoconto della battaglia di Eknomos, dove le forze cartaginesi, tra le quali erano mille frombolieri balearici equipaggiati con armi leggere, sconfissero Agatocle di Siracusa. Ai frombolieri è attribuita la maggior parte del merito della vittoria. Diodoro affermava che i loro proiettili di pietra pesavano una *mina* ciascuno. Ora, la *mina* viene variamente calcolata come l'equivalente di 330 o 540 grammi. Se scegliamo prudentemente il valore inferiore (che corrisponde approssimativamente alla libbra romana e alla *mina* attica) e facciamo l'ipotesi che i proiettili balearici fossero di calcare, ogni pietra avrebbe avuto un diametro di 6,3 centimetri, vale a dire all'incirca le dimensioni di una palla da tennis. Questa dimensione e questo peso costituiscono probabilmente il limite superiore per proiettili da fionda fatti di pietra.

Qual era la precisione dei frombolieri greci e romani e quale danno potevano fare i loro proiettili? Su entrambi i punti si possiedono moltissime testimonianze letterarie. Livio dichiarava che i frombolieri achei erano i migliori: egli attribuiva ciò al fatto che gli achei si allenavano alla precisione scagliando i loro proiettili in modo che passassero attraverso un anello avente una « circonferenza modesta », rizzato a una certa distanza. Grazie a questo addestramento, scrisse Livio, gli achei « potevano colpire non solo il capo dei loro nemici, ma addirittura qualsiasi parte del viso a cui avessero mirato ». Torna cosí subito alla mente il famoso primo tiro di Davide.

Anche l'abilità dei frombolieri balearici veniva attribuita a un addestramento speciale. Diodoro scrive di costoro che « le loro madri li obbligano, ancora fanciulli, a usare la fionda di continuo; infatti di fronte a loro viene posto come bersaglio un pezzo di pane conficcato su un palo, e al novizio non è permesso mangiare finché non abbia colpito il pane, dopo di che lo prende con il permesso della propria madre e lo divora ». Una grande precisione nel tiro con la fionda era attribuita anche ai beniamiti, i frombolieri mancini citati nella Bibbia: essi « lanciavano pietre che potevano colpire un filo di capello senza mancare il colpo ».

Per quanto riguarda l'efficacia della fionda come arma, sarà opportuno far notare che la velocità di un proiettile lanciato con questo strumento può agevolmente superare i cento chilometri orari. Se si fa l'ipotesi che un proiettile di 25 grammi sia dotato di questa velocità quando raggiunge il bersaglio, la sua forza d'urto sarebbe equivalente a quella di una pallina da golf che cadesse dalla cima di una costruzione di sette piani. L'energia di proiettili piú pesanti sarebbe naturalmente maggiore in proporzione. Vegezio diceva che per avversari rivestiti di armature di cuoio i proiettili biconici risultavano piú mortali delle frecce. Anche se il proiettile non penetrava nell'armatura – notava Vegezio – poteva produrre fatali lesioni interne. Ovviamente, se il nemico non possedeva armatura, il proiettile poteva facilmente penetrare nel suo corpo. Celso, l'autore di opere mediche forse migliore di tutto il periodo greco-romano, annoverava nel suo trattato *De medicina* anche talune istruzioni per estrarre proiettili da fionda, di piombo o di pietra, dal corpo dei soldati feriti. Le sue istruzioni seguivano di alcuni secoli l'osservazione dello storico greco Tucidide che i frombolieri di una regione costiera dell'Epiro, gli acarniani, avevano afflitto gli invasori con una tal pioggia di proiettili a lunga gittata che « non era possibile (per gli invasori) muoversi senza armatura ».

Nella storia piú recente, troviamo i commenti dei *conquistadores* sulla precisione e sull'efficacia dei frombolieri peruviani. « La loro arma prin-

I proiettili dei difensori di Olinto durante l'assedio del 348 a.C. avevano un peso che oscillava tra i 19,5 e i 33,4 grammi (*colonne in colore*). Nove dei quattordici tuttavia pesavano meno di 27 grammi. Quando il loro peso fu confrontato con quello degli altri proiettili venuti alla luce durante gli scavi di Olinto (*colonne in grigio*), la maggior parte dei proiettili dei difensori risultò essere del tipo medio-leggero.

I proiettili degli assedianti erano piú pesanti di quelli dei difensori. Dei 23 identificabili come macedoni, 16 oscillavano tra i 30 e i 35,8 grammi. Paragonati al peso degli altri proiettili (*colonne in grigio*), i proiettili macedoni risultano appartenere al tipo pesante.

La distribuzione della fionda nel mondo, dagli albori della storia fino a epoche recenti (*pallini in colore*), rivela che in alcune importanti zone del globo tale arma era sconosciuta. L'importanza della fionda fin dai tempi piú antichi nel Vicino Oriente e in Europa starebbe a dimostrare che questa tecnica si sia diffusa a partire da questo nucleo primitivo. A meno che la fionda non sia stata inventata in modo del tutto indipendente nel Nuovo Mondo, la sua presenza in quel continente potrebbe essere un'altra prova di una qualche forma di comunicazione con il Vecchio Mondo attraverso il Pacifico o la zona subpolare.

cipale », scrisse un osservatore spagnolo, « è la fionda. Con questa essi scagliano una grande pietra con una tale forza da poter uccidere un cavallo. Il suo effetto è in realtà solo un poco inferiore (a quello di un'arma da fuoco spagnola); io ho visto una pietra lanciata da una fionda da una distanza di trenta passi spezzare in due una spada che un uomo teneva in mano ».

Negli anni trenta David M. Robinson, durante i suoi scavi a Olinto, un'antica città della Grecia settentrionale, trovò circa 500 proiettili da fionda di piombo, piú di cento dei quali recavano delle iscrizioni: talune identificano il proiettile come appartenente ai difensori di Olinto o alle truppe macedoni che, guidate da Filippo, padre di Alessandro Magno, conquistarono la città nel 348 a.C. Non tutte le iscrizioni, tuttavia, consentono un'identificazione precisa.

Robinson pubblicò le misure dettagliate di tutti i proiettili che recavano delle iscrizioni. Il peso oscillava tra i 18 e i 35 grammi. Se si confrontano i pesi con le relative iscrizioni, ne emergono curiose correlazioni: la maggior parte dei proiettili macedoni sono tra i piú pesanti, mentre per lo piú quelli olinti sono collocabili tra i leggeri e i medi (*si vedano le illustrazioni a pag. 85*). Questa circostanza fa sí che gli altri proiettili rinvenuti sul luogo possano, almeno in via del tutto teorica, essere attribuiti alle forze macedoni o di Olinto in funzione del loro peso. Inoltre, su alcuni dei proiettili che, per via del loro basso peso, potrebbero essere appartenuti agli olinti compaiono dei nomi: su uno Potalos, su un altro quello che potrebbe essere Timostene o Timostrato. Poiché era usanza diffusa incidere sui proiettili di piombo i nomi dei generali che guidavano l'esercito, è molto probabile che i due uomini, il cui nome è riportato sui proiettili di Olinto, fossero i condottieri delle forze di difesa, altrimenti del tutto sconosciuti. Studi analoghi sui proiettili da fionda provenienti da altri siti d'epoca classica potrebbero fornire informazioni storiche ugualmente sorprendenti.

Le fionde continuarono a essere impiegate per usi militari fino al XVII secolo, ma già nel 400 d.C. l'uso sempre piú diffuso delle armature e della cavalleria veloce rendeva superati i frombolieri. Vegezio raccomandava che essi venissero addestrati a lanciare il proiettile dopo una sola rotazione del braccio invece che dopo tre: lo scopo era naturalmente quello di aumentare il ritmo di lancio del fromboliere. Fu soprattutto il fustibalo che sopravvisse all'introduzione della polvere da sparo e delle armi leggere, ma posso assicurare che l'uso della fionda durò fino al 1936. In quell'anno, durante l'assedio da parte dei lealisti spagnoli dell'Alcazar, la fortezza dove la guarnigione ribelle di Toledo si era rifugiata, gli assedianti si servivano di fionde per tirare le granate all'interno della fortezza. Esiste addirittura un film in cui si può osservare un fromboliere in azione.

Lo studioso inglese di preistoria V. Gordon Childe verso la fine della sua vita piú d'una volta tentò di dimostrare ai propri colleghi l'importanza della fionda come arma bellica. Incontrò scarso successo, ma io, almeno, trovai le sue argomentazioni estremamente interessanti. Infatti, prendendo come spunto il lavoro di Childe, ho pubblicato di recente una ricostruzione ipotetica dell'importanza relativa nelle epoche preistoriche della fionda e dell'arco nel Vicino Oriente. L'area che costituisce l'oggetto del mio studio è delimitata a occidente dal Bosforo, a oriente dall'Indo, a nord dal Caucaso e a sud dal Sinai. In questo vasto territorio, secondo i dati che ho potuto raccogliere, i due tipi di armi si sono rivelati reciprocamente esclusivi per un buon numero di millenni, vale a dire che i popoli che ne usavano una non sembravano scorgere validi motivi per usare anche l'altra.

Questa popolarità della fionda e dell'arco si rivela per la prima volta nel-l'VII millennio a.C. e dura fino al VI millennio o addirittura piú tardi in alcune regioni dell'Asia. Per esempio l'arco era usato quasi esclusivamente in Siria e Palestina prima che in queste zone si organizzassero le città-stato, mentre altre aree del Vicino Oriente preferivano la fionda. Appare evidente che l'arco era rimasto sconosciuto in quelle regioni almeno fino alla fine dell'VIII millennio, mentre l'uso della fionda era di alcune migliaia di anni piú antico. Una sola zona, in cui queste due armi a lunga gittata si sovrappongono, costituisce un'eccezione a tale polarità. Nell'Asia Minore, a Çatal Hüyük e altrove, intorno al 6000 a.C., si usavano sia la fionda sia l'arco.

La mia ipotesi trova conferma nei ritrovamenti relativi a piú di 80 siti della regione, tutti datati in modo abbastanza sicuro, che rivelano la presenza dell'una o dell'altra delle due armi. Questa situazione non è limitata al Vicino Oriente. La presenza dell'arco in Siria e in Palestina rappresenta una specie di testa di ponte asiatica di quest'ultima; l'obiettivo, per cosí dire, doveva essere l'Africa. Qui la preferenza per le frecce è facilmente riscontrabile nelle punte ateriane del Paleolitico superiore, nelle pitture rupestri raffiguranti degli arcieri che sono state scoperte in Africa e, piú lontano, in Spagna, e nelle migliaia di piccole punte di freccia trovate in molti luoghi del Sahara.

Con ogni probabilità la penisola araba può essere considerata come facente parte di questo « regno » di arcieri africani, sebbene la regione sia praticamente terra incognita per gli archeologi. Analogamente, almeno a partire dall'inizio del Neolitico, il regno dei frombolieri dell'Africa sudoccidentale si era ampliato fino ad includere non solo i Balcani, ma anche l'Europa sudorientale in generale. La polarità delle armi a lunga gittata non può essere attribuita a una mancanza di comunicazione tra le due aree, perché i popoli che le abitavano avevano ampi contatti gli uni con gli altri. Per esempio, durante la fase preceramica del Neolitico nel Mediterraneo orientale, l'ossidiana veniva regolarmente trasportata dalle regioni dell'Asia Minore, dove si usava la fionda, fino al lontano sud, come a Beida nel Giordano meridionale, dove l'arco regnava incontrastato. Non è pertanto nell'isolamento che deve essere ricercata una spiegazione. E la spiegazione, una volta trovata, può anche trascendere le armi stesse ed esulare dall'ambito dell'archeologia in senso stretto, che si basa essenzialmente sullo studio di reperti.

Fotosintesi ad alta efficienza

Alcune piante di ambienti caldissimi e aridi hanno sviluppato un insolito processo fotosintetico ad alto rendimento. Questa efficienza potrebbe tornare a vantaggio dell'agricoltura

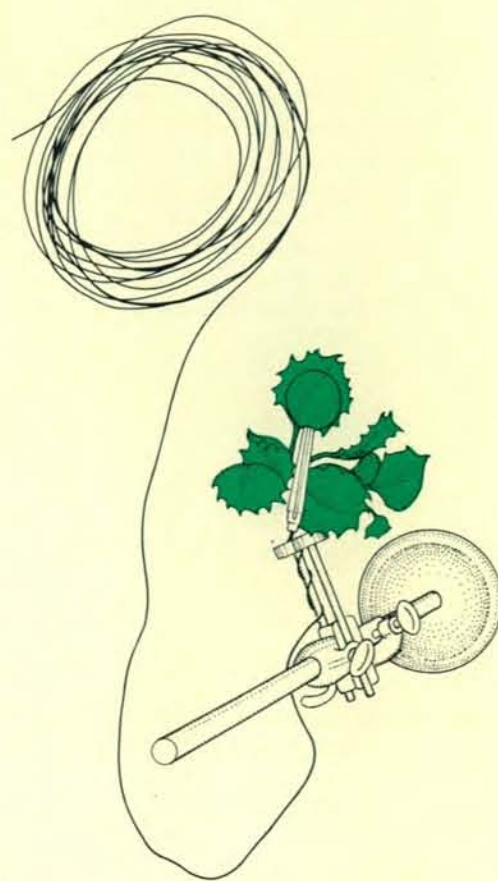
di Olle Björkman e Joseph Berry

Le piante sintetizzano il materiale necessario alla loro crescita dall'acqua, dall'anidride carbonica e dai minerali, attraverso il processo della fotosintesi. Ne consegue che un incremento nell'efficienza della fotosintesi potrebbe portare enormi benefici. I fitofisiologi hanno recentemente appreso che in alcune piante il processo fotosintetico percorre, per quanto riguarda la fissazione dell'anidride carbonica, un unico duplice sistema e che, in alcune condizioni, questo sistema è di gran lunga piú efficiente del processo fotosintetico comune. La scoperta ha sollevato immediate domande. Perché si è evoluto questo processo? Quali generi di piante lo hanno evoluto? Possono acquisirlo le numerose piante che non l'hanno? Le risposte alle prime due domande, almeno, sono completamente chiarite.

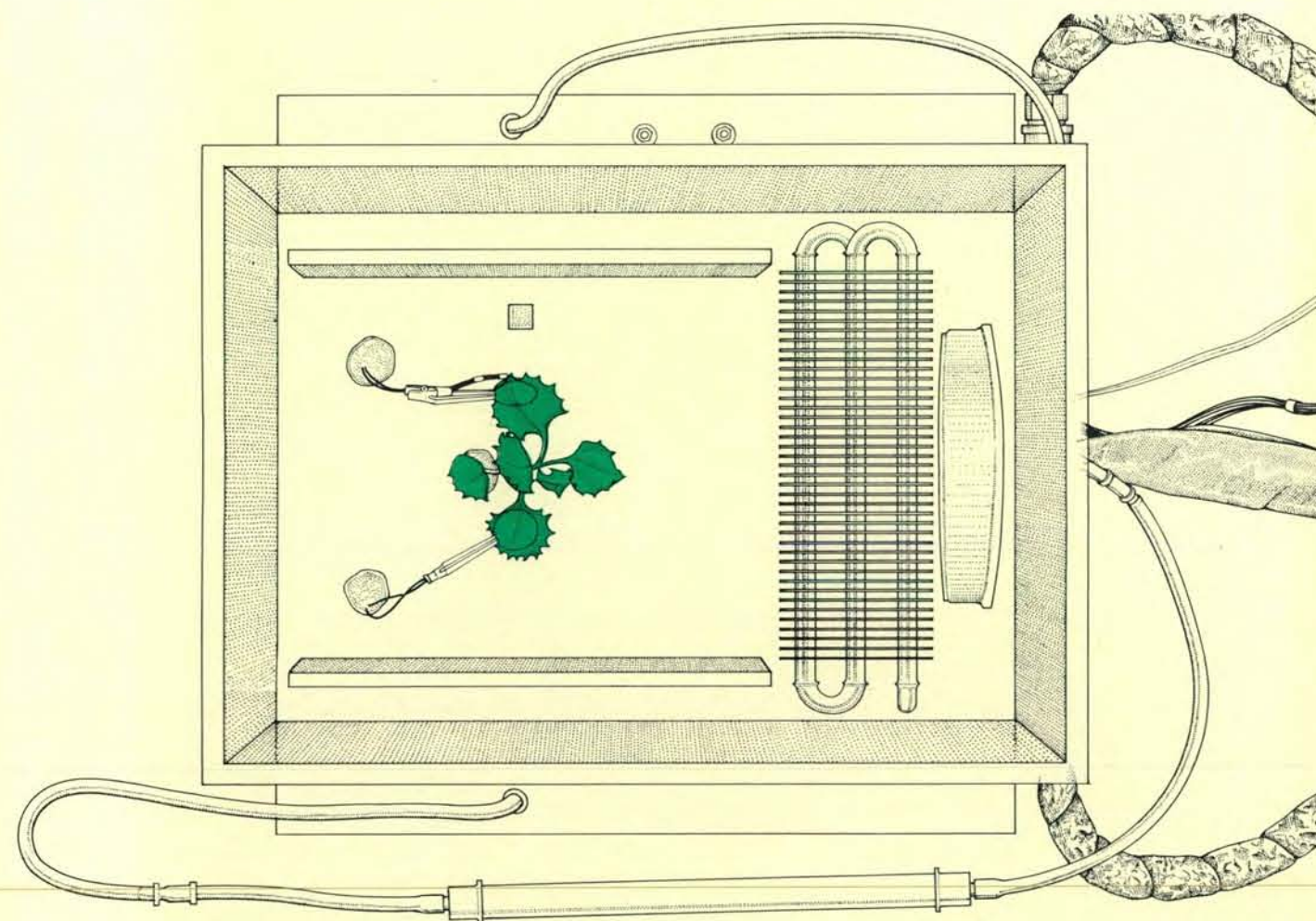
Una vita vegetale, in una forma o nell'altra, può essere trovata in quasi tutti i diversissimi ambienti naturali che esistono sulla Terra. Considerando le limitazioni imposte da cosí diversi habitat, come la tundra artica, l'infuocato deserto o la foresta tropicale piovosa e scura, mi sembra notevole che tanti e diversi generi di piante siano in grado di sopravvivere e riprodursi in ognuno di essi. Succede questo, naturalmente perché le piante stesse sono funzionalmente differenti e sono geneticamente adattate alle condizioni prevalenti nei loro rispettivi habitat. Gran parte dell'adattamento evolutivo è ingegnosamente dovuto alle interazioni tra le piante e i vari componenti dell'ambiente biologico: batteri, altre piante e forme animali, che comprendono protozoi, invertebrati (gli insetti in particolare) e i vertebrati superiori.

Un'altra fonte dell'adattamento è dovuta alle manifestazioni dell'ambiente fisico come, per esempio, la temperatura, il rifornimento idrico e l'intensità luminosa. Queste forme di adattamento interessano, soprattutto, vari processi di crescita delle piante, dei quali il fondamentale è appunto la fotosintesi. La velocità di alcune reazioni fotosintetiche dipende dall'intensità e qualità della luce, la velocità di altre è principalmente influenzata dalla temperatura, e la velocità di altre ancora è in relazione alla concentrazione di anidride carbonica nell'aria. Ne risulta che la velocità del processo fotosintetico, globalmente, è grandemente influenzata dall'ambiente fisico, forse di piú di ogni altro processo accrescitivo. Il rendimento e la produttività di una pianta, sottoposta a stress ambientale, può perciò in gran parte dipendere dall'efficiente funzionamento del suo apparato fotosintetico. Se si sono evolute speciali forme di adattamento che rendono particolarmente efficiente la fotosintesi in condizioni di stress, allora è molto probabile che noi le rinveniamo in piante che vivono in ambienti estremi.

Un sorprendente esempio di tale adattamento è offerto da una pianta erbacea perenne della famiglia delle amarantacee, un gruppo di erbe e arbusti, di cui fa parte la *Tidestromia oblongifolia*, che cresce nelle basse e infuocate aree desertiche a sud-ovest degli Stati Uniti ed è diffusa nella vegetazione del fondovalle della Valle della Morte in California. La Valle della Morte è uno degli habitat piú aspri del mondo, in estate è il piú caldo ambiente naturale di tutto l'emisfero occidentale. Nella Valle della Morte la maggior parte delle piante cresce soltanto nei relativamente tiepidi mesi invernali quando, cioè, cade la maggior parte della scarsa pioggia sul territorio (43 millimetri all'anno). Questo non vale per la *Tidestromia*; quasi tutta la sua attività fotosintetica e il suo sviluppo si manifestano da maggio ad agosto, cioè nei mesi piú caldi e secchi dell'anno.

Con l'aiuto di un laboratorio mobile, noi e i nostri colleghi dell'Istituto Carnegie del Dipartimento di biologia vegetale di Washington e dell'Università di Stanford, abbiamo registrato l'attività fotosintetica delle piante di *Tidestromia* nella Valle della Morte, durante il mese di luglio. Si trovò che le piante fotosintetizzavano ad alta velocità durante le ore diurne. Raggiungevano il massimo a mezzogiorno quando, sia il caldo raggiungeva l'apice, sia le radiazioni solari erano al massimo. (La temperatura ambiente era di circa 50 gradi Celsius). La maggior velocità fotosintetica di queste piante è tra le massime mai registrate in un ambiente naturale ed è paragonabile alla migliore velocità di due piante agricole notevolmente produttive: il granoturco e la canna da zucchero.

Esperimenti durante i quali la temperatura era sotto controllo, mostrarono che la velocità di fotosintesi della *Tidestromia* è grandemente influenzata dai cambiamenti di temperatura. Per esempio, riguardo la velocità di fotosintesi, quando la temperatura è sotto i 20 gradi Celsius, l'arbusto del deserto non eguaglia un'erba adattata alla zona temperata. Quando la temperatura supera i 30 gradi Celsius, invece, la velocità di fotosintesi della pianticella supera quella dell'erba e continua ad aumentare finché raggiunge un picco a 47 °C, temperatura questa che eventualmente sarebbe letale all'erba della zona temperata. Il massimo dell'attività fotosintetica intorno ai 47 °C era stato osservato precedentemente soltanto in una certa alga che si era adattata a un ambiente con primavere caldissime, il che rende davvero fuori del comune il comportamento dell'arbusto desertico. Anche se tali temperature non fossero letali alle piante adattate a vivere in ambienti piú temperati, il meccanismo fotosintetico sarebbe diventato inattivo e la fotosintesi si sarebbe arrestata.

Qual è l'economia idrica di una pianta come la *Tidestromia* che prospera ad alte temperature in un ambiente ad aria estremamente secca e con un rifornimento di acqua molto ridotto? Nelle piante in generale soltanto una piccola frazione dell'acqua assorbita dalle radici serve come reagente nella fotosintesi. La maggior parte traspira come vapore acqueo dalle foglie e si libera nell'aria. Tuttavia la pianta è protetta contro una eccessiva perdita per traspirazione. Il vapore acqueo non viene liberato dalle foglie nell'atmosfera a velocità costante. Le foglie sono ricoperte da una sostanza cerosa impermeabile, e la traspirazione può essere regolata solo attraverso gli stomi che sono pori sulla superficie della foglia. Gli stomi variano la misura della loro apertura; quando sono chiusi, a mala pena poca acqua sfugge dalle umide pareti cellulari della foglia.

Questo fenomeno presenta un enigma. Quando gli stomi chiusi proteggono una pianta da una indesiderata perdita d'acqua, essi impediscono anche all'anidride carbonica atmosferica di entrare all'interno della foglia e di rimpiazzare l'anidride carbonica consumata nel processo fotosintetico. Ne ri-

Le piante del deserto che fioriscono in un habitat caldo e secco sono state saggiate riguardo all'attività fotosintetica in condizioni vicine a quelle naturali, con l'apparecchio raffigurato qui sopra. Le piante sono delle chenopodiacee della specie *Atriplex hymenelytra*. Le termocoppie collegate, a sinistra con la pianta di controllo e a destra con la pianta da esperimento, registravano la temperatura della foglia. Nella camera la temperatura era regolata da un radiatore a circuito idrico e un ventilatore manteneva uniforme la temperatura all'esterno della pianta. La manica d'aria erogava aria di umidità e contenuto di anidride carbonica noti. In queste condizioni la traspirazione d'acqua e l'assunzione di anidride carbonica della pianta da esperimento potevano essere misurate dagli strumenti di laboratorio.

sulta che l'aria dentro la foglia ben presto risulta depauperata di anidride carbonica. Ora, un fattore che limita la velocità di ingresso dell'anidride carbonica nella fotosintesi è la concentrazione dell'anidride carbonica stessa all'interno della foglia. Questo fa sí che il depauperamento di anidride carbonica è in un primo momento rallentato, e poi, quando la concentrazione di anidride carbonica è caduta al di sotto di un certo valore, la fotosintesi comincia ad arrestarsi.

Dal momento che la fuoriuscita di acqua dalla foglia è legata all'ingresso di anidride carbonica, il successo di una pianta in un ambiente caldo e arido dipende in gran parte da quanta anidride carbonica la pianta può fissare fotosinteticamente per unità di acqua perduta nella traspirazione. Il rapporto è noto come « efficienza fotosintetica nell'impiego d'acqua ». L'efficienza, comunque, non è il solo fattore che deve essere tenuto in considerazione. Una riduzione nel diametro degli stomi può incrementare la quantità di carbonio fissato dalle piante rispetto all'acqua perduta e cosí incrementare la sua efficienza nell'impiego dell'acqua. Comunque come abbiamo visto questo tipo di risparmio d'acqua porta inevitabilmente a un abbassamento della velocità assoluta di fotosintesi.

La *Tidestromia*, ha una « efficienza nell'impiego di acqua » molto maggiore di quella che avrebbe una pianta nella zona temperata, se fosse trasferita nell'habitat desertico; nello stesso tempo l'arbusto è in grado di mantenere un'alta velocità fotosintetica. Ricerche di laboratorio hanno dimostrato che l'arbusto può fotosintetizzare a una velocità insolitamente alta anche quando la concentrazione di anidride carbonica dentro le foglie scende a livelli bassi, come accade quando gli stomi sono particolarmente chiusi. Questo mette in grado la pianta di associare un'alta efficienza nell'utilizzazione dell'acqua a un'alta velocità assoluta di fotosintesi. Come si realizza ciò?

La fotosintesi è un processo molto complicato che coinvolge un gran numero di reazioni chimiche. In termini semplicissimi, l'energia luminosa è assorbita dal pigmento della pianta, la clorofilla, e utilizzata per produrre composti intermedi ad alta energia come l'adenosintrifosfato (ATP) e forti riducenti come il nicotinammide-adenindinucleotide ridotto (NADPH) e anche per ossidare l'acqua (una reazione che impiega l'ossigeno gassoso). L'ATP e il NADPH guidano reazioni che utilizzano l'anidride carbonica atmosferica e la riducono per formare carboidrati, amminoacidi e altri costituenti della pianta.

Esaminiamo ora piú dettagliatamente l'ultima parte del processo. In quasi tutte le piante, l'anidride carbonica dell'aria reagisce con un composto, che si è formato all'interno della foglia: il ribulosio-1,5-difosfato, o RuDP, zucchero fosforilato a 5 atomi di carbonio. Quando una molecola di anidride carbonica reagisce con una molecola di RuDP (la reazione è catalizzata dall'enzima RuDP-carbossilasi), si formano due molecole a tre atomi di carbonio ognuna. La sostanza a tre atomi di carbonio appena formata è l'acido fosfoglicerico (PGA). La maggior parte dell'acido fosfoglicerico formatosi dentro la foglia è convertito in vari prodotti finali (come il saccarosio) attraverso una serie di reazioni endoenergetiche. Una parte dell'acido fosfoglicerico, comunque, serve per rigenerare le molecole di RuDP, cosicché esse possono di nuovo reagire come accettori di anidride carbonica. Questa rigenerazione completa un ciclo che rende il processo di fissazione dell'anidride carbonica autosufficiente. Il processo principale è chiamato ciclo di Calvin-Benson dai suoi scopritori, Melvin Calvin e Andrew A. Benson.

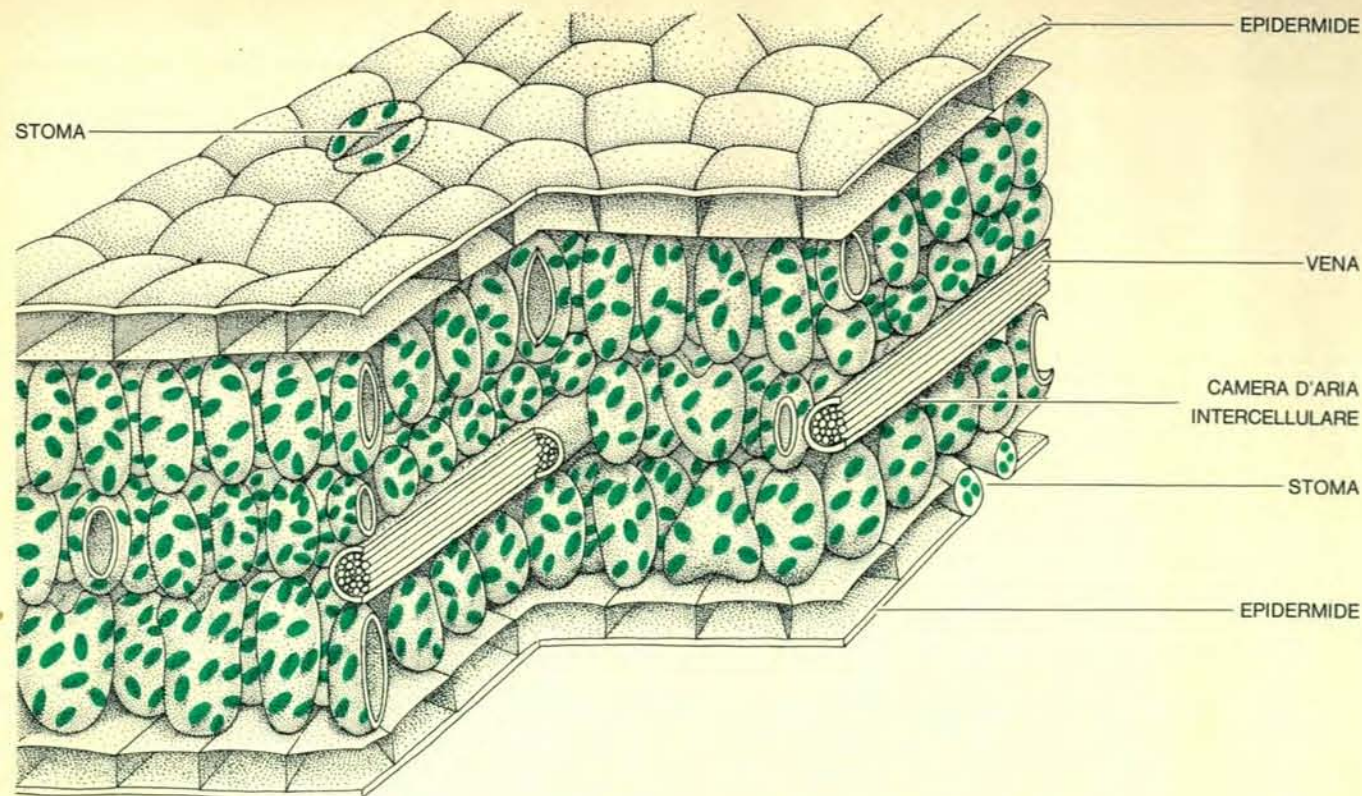
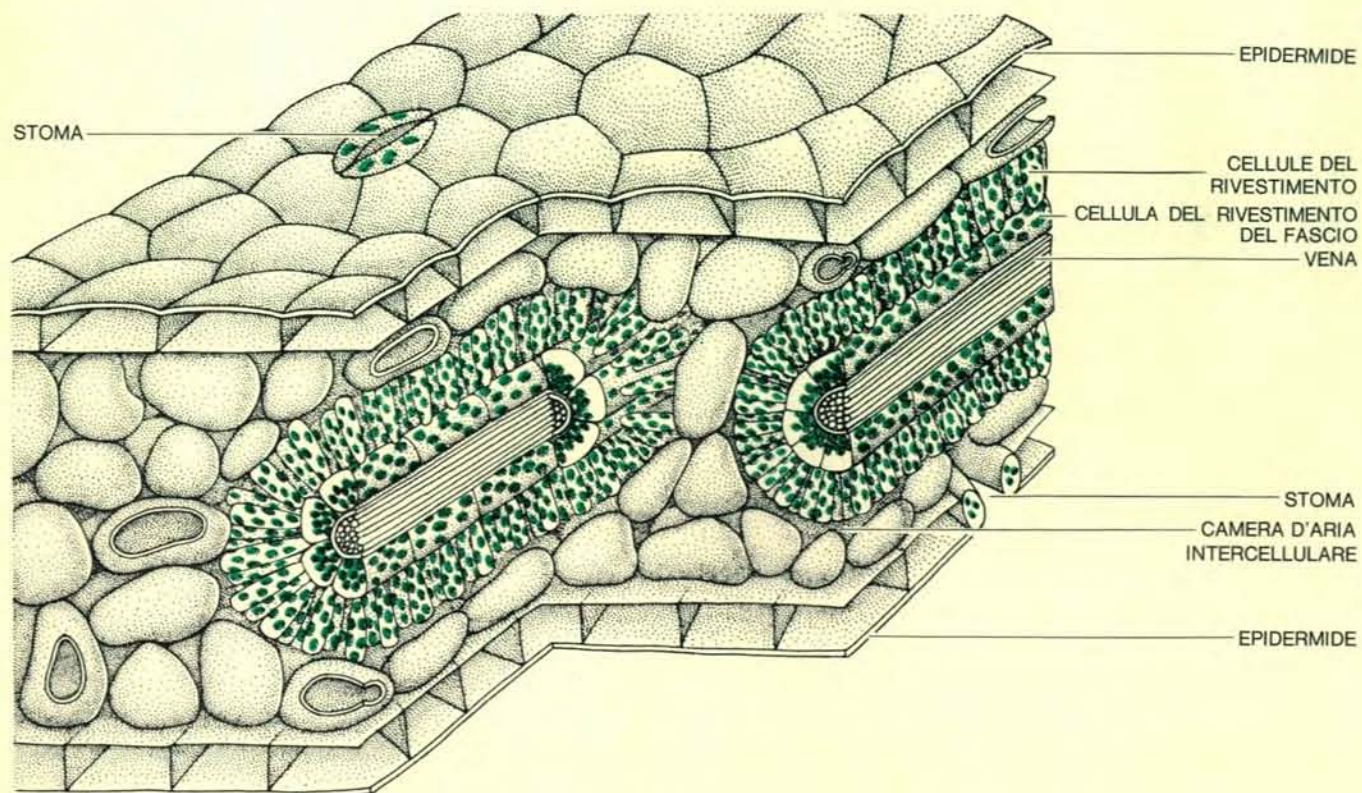
In poche piante, e tra queste l'arbusto desertico *Tidestromia*, la fissazione iniziale dell'anidride carbonica atmosferica è realizzata in modo abbastanza diverso. In queste piante le molecole accettrici all'interno delle foglie sono quelle di fosfoenolpiruvato (PEP). Quando una molecola di anidride carbonica reagisce con una molecola di PEP (la reazione è catalizzata dall'enzima PEP-carbossilasi), il risultato finale è la formazione di acido malico e acido aspartico. In ognuna di queste molecole ci sono quattro atomi di carbonio (C_4), le differenze tra le attività del PEP e del RuDP e i loro enzimi sono molto significative. In primo luogo, il PEP è molto piú reattivo con l'anidride carbonica del RuDP. In secondo luogo, l'attività del RuDP-carbossilasi è in qualche modo inibita dall'ossigeno, invece quella del PEP carbossilasi non lo è. Ne consegue che la fissazione di anidride carbonica nel sistema del PEP è molto piú efficace di quella del sistema del RuDP, quando la concentrazione di anidride carbonica nell'atmosfera è bassa e quella dell'ossigeno alta.

La fissazione dell'anidride carbonica atmosferica da parte delle molecole accettrici di PEP è, comunque, soltanto l'inizio del processo in piante come la *Tidestromia.* Nè l'uno nè l'altro dei prodotti di questo ciclo (acido molico e acido aspartico) può compiere funzioni equivalenti a quelle svolte dall'acido fosfoglicerico nel ciclo di Calvin-Benson; questi acidi non possono essere trasformati in carboidrati o in altri prodotti finali utili nella fotosintesi, senza perdita di carbonio. In seguito accade che l'acido malico e l'acido aspartico siano scissi enzimaticamente; questo processo da una parte rilascia l'anidride carbonica precedentemente fissata e dall'altra parte genera molecole di acido piruvico a tre atomi di carbonio. Il passaggio successivo chiude il ciclo dell'acido piruvico e lo rende autosufficiente; infatti l'acido piruvico reagisce con l'ATP generato dalla fotosintesi per formare le molecole di fosfoenolpiruvato (PEP), che sono libere di agire come accettori di anidride carbonica.

In conclusione il sistema del PEP fissa l'anidride carbonica solo per rilasciarla di nuovo. Questa poteva sembrare essere una attività senza scopo finché si riuscí a capire che il sito di fissazione era a una distanza fisica significativa dal sito di rilascio. Inoltre, appena rilasciata, l'anidride carbonica è immediatamente utilizzata dal sistema del RuDP della pianta ed entra nel ciclo fotosintetico di Calvin-Benson. Nel-

ARIA ESTERNA
ANIDRIDE CARBONICA
STOMA (APERTO)
EPIDERMIDE
CAMERA D'ARIA INTERCELLULARE
PARETE CELLULARE
RuDP
ATP
RuDP CARBOSSILASI
CARBOIDRATI
CICLO DI CALVIN-BENSON
CELLULA FOTOSINTETICA
PGA
ATP
NADPH

Il sistema fotosintetico « C_3 » è quello normale, così chiamato perché il prodotto iniziale è un composto con tre atomi di carbonio per molecola. In questo diagramma schematico e semplificato, l'anidride carbonica che la foglia assorbe dall'aria circostante reagisce con il ribulosiodifosfato (RuDP) in una reazione catalizzata dall'enzima RuDP-carbossilasi, formando due molecole di acido fosfoglicerico (PGA) a tre atomi di carbonio ciascuna. Nelle reazioni successive una parte dell'acido fosfoglicerico è trasformata in prodotti finali della fotosintesi, e una parte viene utilizzata per rigenerare molecole di RuDP cosicché esse possono di nuovo funzionare da accettori di anidride carbonica. Il cerchio così si chiude e il processo di fissazione di anidride carbonica risulta un ciclo autosufficiente guidato dall'energia luminosa. I prodotti finali della fotosintesi sono i carboidrati, gli amminoacidi e altri composti di cui la pianta ha bisogno per la crescita e la sopravvivenza.

Il sistema fotosintetico « C_4 » si manifesta in alcune piante del deserto specificamente adattate. L'anidride carbonica che entra nella foglia reagisce col fosfoenolpiruvato (PEP), un composto a tre atomi di carbonio per formare acido ossalacetico (OAA) a quattro atomi di carbonio, dal quale poi si forma acido malico e acido aspartico. Questi sono trasportati dalle cellule del mesofillo, che sono nella parte più esterna della foglia, alle cellule più interne del rivestimento del fascio. Qui l'anidride carbonica è rilasciata dai prodotti « C_4 » e si forma acido piruvico, composto a tre atomi di carbonio. L'anidride carbonica è ora fissata di nuovo nel ciclo normale. L'acido piruvico torna alle cellule del mesofillo e acquista un gruppo fosfato dall'adenosintrifosfato (ATP) per formare PEP, rigenerando così l'iniziale molecola accettrice di anidride carbonica. Questo ciclo addizionale per la fissazione di anidride carbonica contribuisce a incrementare l'efficienza globale della sua utilizzazione.

Struttura della foglia di una pianta « C_3 » di *A. patula*. Come in altre foglie tipiche, le cellule che contengono clorofilla (*in colore*) sono di un solo tipo e sono distribuite in tutto l'interno della foglia. *A. patula* ha un'affine, *A. rosea*, che utilizza per la fotosintesi il sistema « C_4 » e ha la struttura della foglia diversa, come è mostrato dall'illustrazione sotto.

La foglia di *A. rosea* ha quasi tutta la clorofilla in due tipi di cellule che formano due cilindri concentrici. Le cellule del cilindro esterno sono le « cellule del mesofillo »; quelle del cilindro interno sono le « cellule del rivestimento del fascio ».

le piante come la *Tidestromia*, allora, il sistema del PEP fornisce un mezzo iniziale e molto efficiente per fissare l'anidride carbonica atmosferica, anche a concentrazioni molto basse. Il sistema serve a un unico scopo: fornire l'anidride carbonica al ciclo fotosintetico di Calvin-Benson. Dal momento che il primo prodotto del ciclo di Calvin-Benson è una sostanza a tre atomi di carbonio per molecola, le piante che effettuano la fotosintesi soltanto per mezzo di questo ciclo sono ora comunemente chiamate « piante C_3 » o piante a tre atomi di carbonio. Questo le distingue dalle piante del tipo della *Tidestromia*, che fissano l'anidride carbonica atmosferica per mezzo del sistema del PEP; queste sono conosciute come « piante C_4 » o piante a quattro atomi di carbonio, poichè i prodotti del sistema del PEP, acido malico e acido aspartico, hanno quattro atomi di carbonio per molecola.

La prima prova dell'esistenza di piante a quattro atomi di carbonio fu ottenuta all'inizio del 1960 da Hugo Kortschak e dai suoi colleghi nel Laboratorio dell'Associazione dei proprietari di piantagioni di zucchero hawaiani. Essi fornirono anidride carbonica marcata con carbonio-14 alle piante di canna da zucchero che fotosintetizzavano e pochi secondi dopo analizzavano i composti delle loro foglie. Essi trovarono che gli atomi di carbonio-14 erano soprattutto concentrati in un composto a 4 atomi di carbonio, l'acido malico.

La scoperta era in diretta contraddizione con i risultati di esperimenti similari che facevano uso di piante che andavano dalle alghe primitive alle piú avanzate angiosperme.

In tutti questi esperimenti l'anidride carbonica marcata era per prima fissata nel composto a tre atomi di carbonio, l'acido fosfoglicerico. Sotto alcuni punti di vista questo processo, scoperto da Kortschak e dai suoi collaboratori, assomiglia a quello usato per la fissazione dell'anidride carbonica da alcune piante dolci. Anche batteri e animali hanno l'enzima PEP-carbossilasi, ma il suo livello di attività è molto piú basso e le sue funzioni sono differenti da quelle svolte nelle « piante C_4 ». Subito dopo la scoperta di Kortschak, due fitofisiologi australiani, Hal Hatch e Roger Slack, confermarono i suoi risultati e in una serie di eleganti esperimenti chiarirono le principali reazioni del « sistema C_4 ». Non ci volle molto tempo perché la lista delle piante conosciute come « C_4 » venisse ampliata; giunse a comprendere non soltanto la canna da zucchero, ma anche altre piante da raccolta della famiglia delle graminacee (come il granoturco e il sorgo) e molte erbe estive (come la sanguinaria e la gramigna).

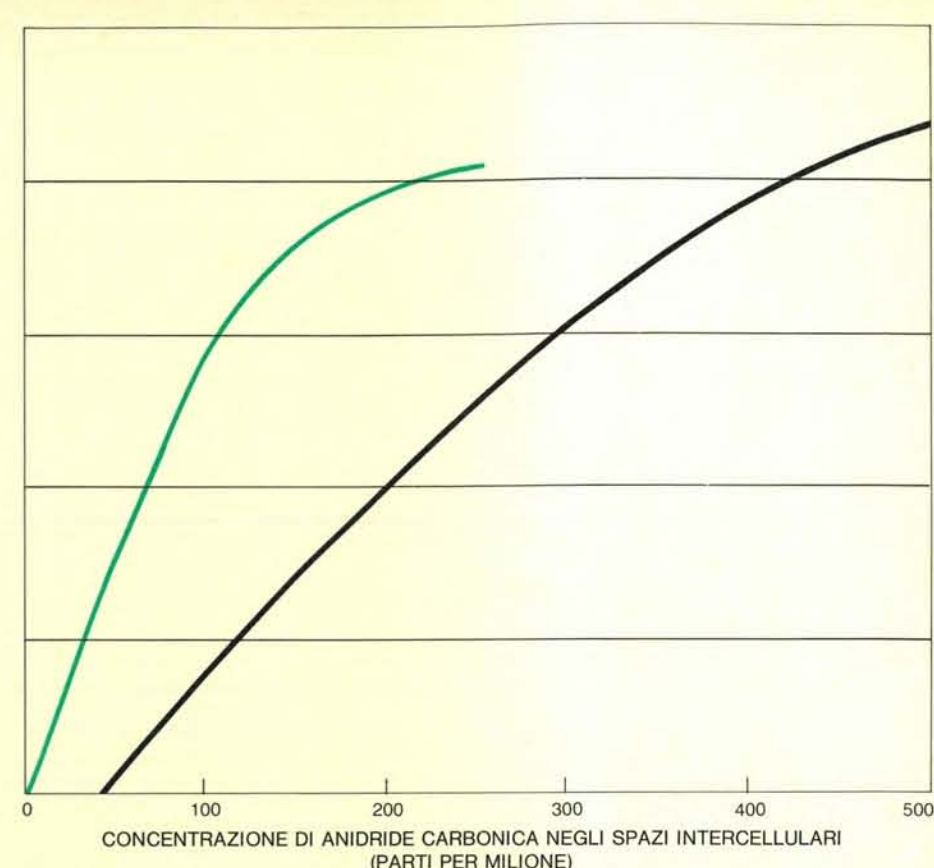
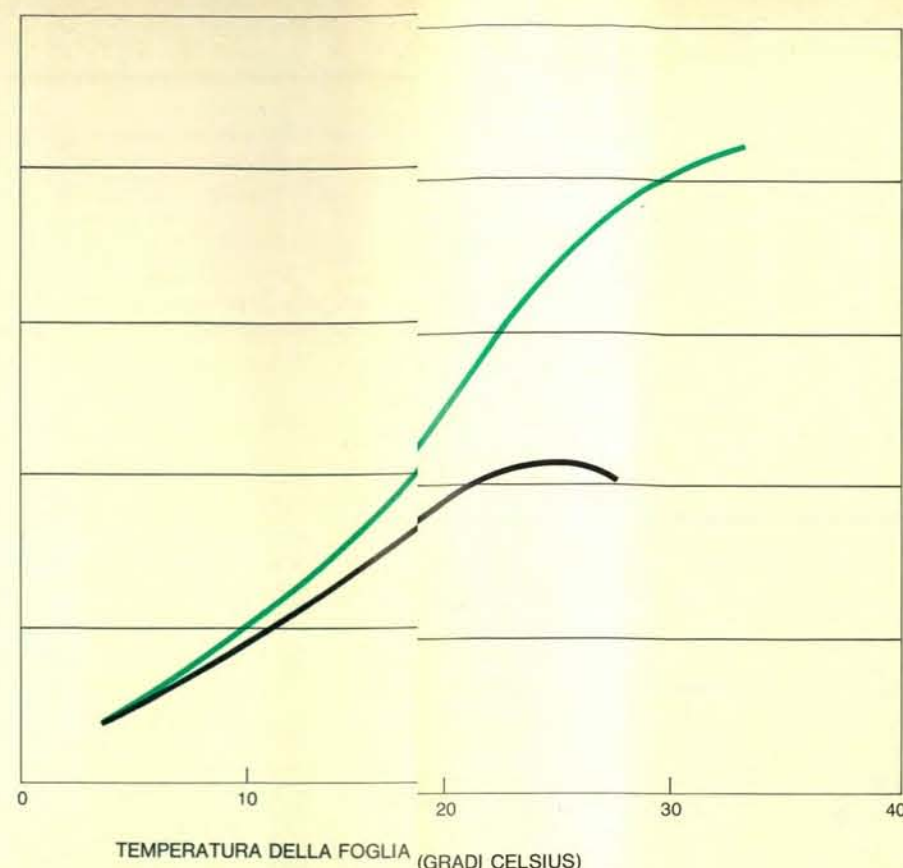
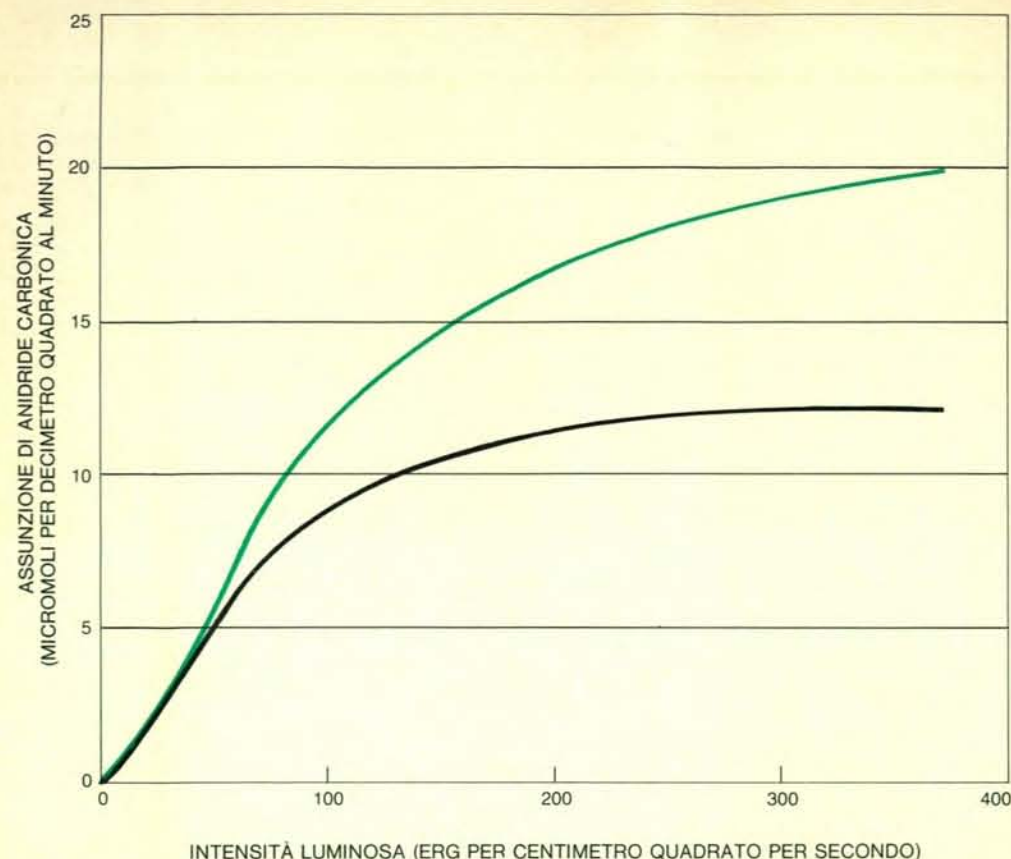
Oggi si sa che il processo « C_4 » esiste in quasi cento generi, in almeno 10 famiglie vegetali, sia monocotiledoni che dicotiledoni; e il numero delle specie a quattro atomi di carbonio è dell'ordine delle centinaia. Cosí non sorprendentemente si trovò che l'arbusto del deserto *Tidestromia* era una pianta a quattro atomi di carbonio. Cosí parecchie altre piante della Valle della Morte del genere *Atriplex*. Questo genere è uno degli undici generi che comprendono specie vegetali del tipo a tre atomi di carbonio e a quattro atomi di carbonio. Considerando che il processo « C_4 » può essere presente in alcune specie e non in altre, e che le piante « C_4 » sono state trovate in famiglie vegetali cosí poco correlate come le erbe e i girasoli, una considerazione appare inevitabile: il processo a quattro atomi di carbonio deve essersi evoluto abbastanza indipendentemente un certo numero di volte.

Ben presto si convenne che le piante « C_4 », qualunque fossero le loro relazioni tassonomiche reciproche, condividevano alcune caratteristiche addizionali che le distinguevano dalle piante « C_3 ». Una di queste caratteristiche, sulla quale ritorneremo, è una struttura anatomica specializzata della foglia, che ha senso solo per il particolare primo passaggio, proprio delle fotosintesi « C_4 ». Un'altra è la presenza di alte concentrazioni di alcuni enzimi nelle foglie delle piante a quattro atomi di carbonio, che invece si trovano a concentrazioni molto piú basse nelle foglie delle piante a tre atomi di carbonio. Come potrebbe essere anticipato, questi enzimi giocano un ruolo importante nelle sintesi del processo « C_4 », ma non sono coinvolti nel processo « C_3 ».

Una terza caratteristica, risultato della diversità tra i modi di fissazione dell'anidride carbonica delle piante « C_3 » e « C_4 », è la differenza nelle proporzioni dei due isotopi naturali non radioattivi del carbonio (carbonio-12 e carbonio-13) presenti nei composti del carbonio dei due gruppi di piante. Questa differenza, tra parentesi, può essere usata per distinguere lo zucchero della barbabietola da zucchero (una pianta a quattro atomi di carbonio) e lo zucchero della canna da zucchero. Il saccarosio puro delle due piante è uguale sotto tutti i punti di vista tranne che per il rapporto dei due carboni radioattivi.

E veniamo ora alla domanda del motivo per cui le piante « C_4 » si sono evolute. Questa domanda sorse spontanea in seguito a osservazioni sulle piante « C_4 » stesse, sia in natura che in laboratorio. Per esempio, se una pianta « C_4 » è collocata in un ambiente chiuso, cosicché l'aria possa essere prelevata per l'analisi, si nota che l'attività fotosintetica della pianta depaupera quasi interamente l'aria di anidride carbonica. Per un paragone, una pianta « C_3 » in un ambiente chiuso simile può assorbire anidride carbonica finché la concentrazione scende dal livello normale di 300 parti per milione fino a circa 50 parti per milione e se la concentrazione scende ancora piú in basso di 50 parti per milione, la pianta « C_3 » non assorbe ulteriormente anidride carbonica ma anzi ne emette.

Prendiamo ora in considerazione un esperimento impostato sul collocare in uno stesso ambiente chiuso una pianta « C_4 » e una pianta « C_3 ». Quando l'attività fotosintetica di entrambe le piante ha ridotto la concentrazione di anidride carbonica a meno di 50 parti per milione, la pianta « C_3 » può cominciare a emettere anidride carbonica. La pianta « C_4 » ora può continuare ad assorbire la scarsa provvista di anidride carbonica e mantenersi in vita a spese della pianta « C_3 », finché, impoverita della scorta di carbonio, la pianta a tre atomi di carbonio muore. La capacità delle piante « C_4 » di assorbire anidride carbonica dall'aria, anche quando la concentrazione è cosí bassa da aggirarsi intorno a una o due parti per milione, ha fornito un'utile criterio di esame per determinare se le varie specie di piante possedessero o no il sistema « C_4 ».

La capacità di eliminare dall'aria l'anidride carbonica e di abbassare il rapporto del carbonio-13 rispetto al carbonio-12 riflette alcune importanti caratteristiche intrinseche del sistema « C_4 ». In circostanze normali, comunque, noi non ci aspetteremmo che né una né l'altra di queste caratteristiche possano di per sé avvantaggiare selettivamente le piante a quattro atomi di carbonio. Il vantaggio conseguito dalle piante « C_4 », cioè una piú elevata velocità massima di fotosintesi, può essere ottenuto soltanto quando la luce è intensa e la temperatura assai alta. Per esempio, i generi vegetali « C_4 » come la *Tidestromia*, sono rinvenuti soprattutto in zone in cui la stagione del piú attivo sviluppo delle piante coincide con il periodo di massima radiazione solare, di massima temperatura e di piú limitato rifornimento di acqua. Contemporaneamente, comunque, anche un certo numero di piante « C_3 »

L'andamento della fotosintesi delle piante strettamente affini di *A. patula* (*nero*) e di *A. rosea* (*colore*) varia considerevolmente anche quando le piante crescono in condizioni di controllo identiche. *A. patula* utilizza il sistema fotosintetico « C_3 »; *A. rosea* il sistema « C_4 ». Le differenze aumentano con l'aumentare dell'intensità luminosa e della temperatura della foglia e col diminuire dell'anidride carbonica. Poiché le piante sono affini, la differenza riflette la diversità di fotosintesi dovuta a cicli diversi.

è presente in questi aspri habitat e un certo numero di piante « C_4 » cresce rigogliosamente in ambienti molto piú temperati.

Teoricamente, per scoprire quale è il significato funzionale e adattativo del sistema « C_4 », lo sperimentatore dovrebbe avere a sua disposizione due gruppi di piante identiche nel corredo genetico, eccetto che per dei geni specifici, responsabili, in un gruppo, del processo « C_4 ». E succede che una situazione abbastanza vicina a questa teorica si verifichi con le specie « C_3 » e « C_4 » di *Atriplex*. Alcuni elementi di questa specie sono geneticamente abbastanza simili, tanto da ibridizzare, e sono dotati di un alto grado di appaiamento cromosomico. Inoltre, due di queste (la specie « C_4 » *A. rosea* e la specie « C_3 » *A. patula*) coesistono in uno stesso ambiente e hanno procedimenti simili di crescita estiva.

Un modo per valutare il valore adattativo del processo a quattro atomi di carbonio è confrontare l'efficienza fotosintetica di una tale coppia di piante quando si varino le componenti principali dell'ambiente fisico. Le risposte delle piante ai cambiamenti insieme con le conoscenze di come mutino le stesse componenti fisiche in natura, indicheranno le efficienze relative delle due piante in un vasto campo di habitat potenziali.

Noi abbiamo raccolto risultati di questo tipo, seguendo la crescita di *A. rosea* e di *A. patula* in laboratorio, sottoponendole a identiche condizioni controllate, incluse le condizioni quasi ottimali di ognuna delle due specie (*si veda l'illustrazione in alto*). Abbiamo poi paragonato la loro velocità fotosintetica (espressa in unità basate sull'area superficiale delle foglie delle piante) in un'atmosfera normale al variare dell'intensità luminosa, e abbiamo trovato che a 25 °C e a basso livello di illuminazione, le specie « C_3 » e « C_4 » fotosintetizzavano a velocità molto simili. Aumentando l'intensità luminosa, anche entrambe le piante aumentavano la velocità di fotosintesi. Alla massima intensità solare, la velocità di fotosintesi della *A. rosea* (C_4) era molto piú elevata di quella della *A. patula* (C_3). Abbiamo concluso che le piante « C_4 » possono essere piú efficienti delle « C_3 » in habitat sereni e soleggiati, ma che se la luce solare è offuscata dall'ombra delle nuvole o di altre piante, la superiorità di una specie sull'altra può essere relativamente insignificante.

Poi abbiamo variato la temperatura, lasciando costante il livello di illuminazione. Abbiamo scoperto che la velocità diminuiva e diventava addirittura insignificante alla temperatura di 5 °C. Aumentando la temperatura oltre i 25 °C, avevamo l'effetto opposto: a 33 °C, la velocità fotosintetica della *A. rosea* (C_4) era piú che doppia di quella della *A. patula* (C_3).

I risultati degli esperimenti riguardo la temperatura e l'illuminazione ci hanno condotto a due conclusioni. In primo luogo la specie di *Atriplex* a quattro atomi di carbonio è in grado di sintetizzare piú vantaggiosamente in condizioni associate di intense radiazioni solari e di temperatura elevata caratteristiche di un habitat arido a giorni caldi e sereni. In secondo luogo questa superiorità della *A. rosea* rispetto alla sua strettamente affine (dal punto di vista genetico) *A. patula*, ci porta a credere che la maggior efficienza della pianta « C_4 » sia in gran parte attribuibile al fatto che la pianta possiede il processo « C_4 » e non ad altri possibili adattamenti, svincolati dal processo « C_4 », e probabilmente evolutisi contemporaneamente sotto la pressione della selezione naturale.

Noi abbiamo già sottolineato la relazione tra l'efficienza di una pianta nell'utilizzare acqua e l'apertura degli stomi delle sue foglie, e anche l'effetto inibente degli stomi chiusi sull'assunzione di anidride carbonica. È evidente che un motivo della superiorità delle piante « C_4 » è la loro capacità a continuare la fotosintesi anche quando l'anidride carbonica presente nell'aria degli spazi intercellulari della foglia è in concentrazioni bassissime. Considerando gli stomi della foglia completamente aperti e un'atmosfera contenente la concentrazione normale di anidride carbonica di 300 parti per milione, si calcola che all'interno della foglia di una *A. rosea* (C_4), la concentrazione di anidride carbonica possa essere di circa 195 parti per milione, e invece all'interno di una foglia di *A. patula* (C_3) possa essere di 240 parti per milione. Uno comunque non dovrebbe essere sviato dalla maggiore concentrazione rinvenuta nelle foglie della pianta « C_3 ». Le 240 parti per milione rappresentano meno della metà della concentrazione di anidride carbonica necessaria per saturare il processo fotosintetico di una pianta « C_3 »; mentre invece 195 parti per milione saturano quasi del tutto il processo fotosintetico della pianta « C_4 ».

Naturalmente, quando gli stomi delle foglie variano da particolarmente chiusi a completamente aperti, la concentrazione intercellulare dell'anidride carbonica nelle foglie viene ridotta proporzionalmente. Non appena la concentrazione cade al di sotto di 200 parti per milione, *A. rosea* comincia a superare *A. patula* nell'efficienza di fotosintesi. C'è comunque, un altro lato della medaglia. Se la concentrazione di anidride carbonica è mantenuta a livelli anormalmente alti o anormalmente bassi, *A. patula* allora eguaglia, se non supera, l'efficienza fotosintetica della *A. rosea*.

Ora, la velocità di traspirazione a un dato diametro degli stomi (o, piú precisamente, la conduttanza degli stomi) è la stessa in entrambe le piante « C_3 » e « C_4 » poiché la velocità è determinata esclusivamente dalla differenza di concentrazione di vapore acqueo tra l'interno e l'esterno della foglia. Perché, dunque, c'è una differenza nell'efficienza di utilizzazione d'acqua tra le due specie di piante? L'efficienza fotosintetica del processo « C_4 » fornisce la risposta. Dal momento che il processo « C_4 » può operare a concentrazioni intercellulari molto basse di anidride carbonica, una pianta « C_4 » mantiene una differenza maggiore tra le concentrazioni di anidride carbonica dell'ambiente esterno e dell'interno della foglia. A un dato diametro degli stomi, quindi, la diffusione di anidride carbonica verso l'interno della foglia è un processo piú veloce per una pianta « C_4 » che per una « C_3 ». Il risultato conseguente è una piú alta velocità di fotosintesi per ogni data velocità di perdita di vapore acqueo. Per esempio, possiamo calcolare la conseguenza provocata dalla riduzione dell'apertura degli stomi (o della conduttanza) delle foglie di *A. rosea* finché la velocità di fotosintesi della pianta non superi piú la velocità massima della *A. patula*. In queste condizioni, la pianta « C_4 » traspira soltanto un quinto dell'acqua traspirata dalla pianta « C_3 ».

Noi abbiamo cercato di sperimentare queste previsioni con test di crescita comparata nel giardino sperimentale dell'Istituto Carnagie nell'Università di Standford. Il luogo, la Valle di Santa Clara, ha un clima mediterraneo. Le estati sono calde e poco piovose e, se le piante del giardino non vengono irrigate, il rifornimento idrico disponibile è limitato all'acqua che era presente nel suolo all'inizio della stagione della crescita.

Noi abbiamo piantato in marzo in due « aree di esame » pianticelle di entrambe le specie *A. rosea* e *A. patula* e abbiamo bagnato le pianticelle uniformemente fino all'inizio di maggio.

Dopo questa data abbiamo irrigato un'area ma non l'altra. Entrambe le specie nell'area irrigata crescevano a circa la stessa velocità, maturavano bene e producevano semi in abbondanza. Nell'area non irrigata la pianticella di *A. rosea* cresceva a velocità molto simile a quella delle pianticelle dell'area irrigata, ma la crescita delle pianticelle di *A. patula* rallentava bruscamente quando l'irrigazione cessava. Tutte le piante « C_3 » morivano prima di raggiungere la maturità, mentre le piante « C_4 » crescevano rigogliosamente, maturavano e producevano semi.

Un ulteriore fattore che influenza l'efficienza del processo metabolico vegetale sembra essere di antica origine nel corso dell'evoluzione. Durante la maggior parte del periodo Precambriano, prima che si siano evolute le piante fotosintetiche, l'atmosfera terrestre conteneva solo circa l'1 per cento di ossigeno. Oggi, centinaia di milioni di anni dopo, l'atmosfera contiene circa il 21 per cento di ossigeno, e questo è il risultato della fotosintesi vegetale. Attualmente, la nostra alta concentrazione di ossigeno nell'atmosfera inibisce la fotosintesi per quanto riguarda le piante « C_4 ». L'effetto è relativamente trascurabile quando la temperatura è bassa, ma diventa determinante quando la temperatura aumenta. Se una pianta « C_3 » è collocata in una atmosfera artificiale che contiene la normale concentrazione di anidride carbonica dello 0,03 per cento, ma una concentrazione di ossigeno solo dell'1,5 per cento, la sua velocità di assunzione di anidride carbonica è circa il 40 per cento piú alta di quella che si verifica in una normale atmosfera con il 21 per cento di concentrazione di ossigeno. Viceversa, se la concentrazione di ossigeno è portata al di sopra del 21 per cento e la concentrazione di anidride carbonica abbassata a livelli inferiori al valore normale dello 0,03 per cento, la velocità di assunzione di anidride carbonica delle piante « C_3 » cade in maniera concorde. L'effetto inibente, comunque, è completamente eliminato, se la concentrazione di anidride carbonica è portata a livelli abbastanza alti da saturare il processo fotosintetico.

ATRIPLEX ROSEA

ATRIPLEX PATULA

IBRIDI DI SECONDA GENERAZIONE

L'anatomia delle foglie di *A. rosea*, di *A. patula* e di un ibrido ottenuto dall'incrocio delle due piante, è mostrata in alto in sezione trasversale. Nell'ibrido di prima generazione l'anatomia della foglia associa i caratteri di entrambi i tipi parentali. Negli ibridi di seconda generazione l'anatomia varia molto da pianta a pianta. Alcuni hanno distinte cellule del mesofillo e cellule del rivestimento del fascio come nel genitore *A. rosea*. Altri mancano di questa specializzazione come in *A. patula*.

L'effetto inibente dell'ossigeno atmosferico non è apprezzabile, per quanto riguarda le piante « C_4 », anche quando la concentrazione di anidride carbonica cade al di sotto del valore normale dello 0,03 per cento. Poiché una atmosfera povera di ossigeno incrementa la velocità di fissazione dell'anidride carbonica in una pianta « C_3 », sembra logico aspettarsi che la differenza nel compimento della fotosintesi tra le piante « C_3 » e le piante « C_4 » diventerà insignificante in tale atmosfera. E gli esperimenti lo confermano: quando *A. rosea* e *A. patula* sono collocate in un ambiente chiuso con aria che contiene soltanto l'1,5 per cento di ossigeno senza tener conto della temperatura e della illuminazione, le velocità di fotosintesi delle due specie sono molto simili. Lo stesso vale in una atmosfera di concentrazione normale d'ossigeno, purché la concentrazione di anidride carbonica sia stata portata al punto di saturazione. In definitiva, gli studi comparativi su *A. rosea* e *A. patula*, comprese le scoperte riguardo all'effetto inibitorio dell'ossigeno sulla fotosintesi delle piante « C_3 », conducono alla seguente conclusione. La particolare proficuità della attività fotosintetica delle piante « C_4 » in condizioni di intense radiazioni, alte temperature e limitato rifornimento d'acqua, è in gran parte il risultato di una maggiore efficienza della pianta nell'utilizzazione dell'anidride carbonica a basse concentrazioni. Come è stato visto questa efficienza è attribuibile soprattutto al ciclo « C_4 » per la fissazione del carbonio. Il ciclo « C_4 » comunque non è l'unico fattore in giuoco; anche l'anatomia della foglia delle piante « C_4 » svolge un ruolo significativo. Confrontiamo brevemente le foglie di *A. patula* e *A. rosea*.

La fotosintesi e la traspirazione di *A. rosea* e di *A. patula* sono raffigurate qui sopra. La traspirazione d'acqua dalla foglia (*in nero*) è la stessa per entrambe e diminuisce con la conduttanza degli stomi. La fotosintesi diminuisce al chiudersi degli stomi. A una data conduttanza stomatica, *A. rosea* (*linea continua colorata*) è superiore all'affine « C_3 » *A. patula* (*linea spezzata colorata*) per quanto riguarda l'assunzione di anidride carbonica. A una data velocità di assunzione di anidride carbonica la pianta « C_4 » può tenere i suoi stomi meno aperti e cosí può perdere meno acqua della pianta « C_3 ».

L'anatomia delle foglie di *A. patula* è quella comune alle piante « C_3 ». Le cellule che contengono cloroplasti (organelli citoplasmatici, in cui l'energia luminosa viene assorbita dalla clorofilla) sono distribuite in tutta la foglia. Sistemi identici di enzimi fotosintetici sono presenti in ogni cellula che contiene cloroplasti. Ogni cellula, che segue il ciclo di Calvin-Benson, fissa, indipendentemente dalle altre, una piccola quantità di anidride carbonica dell'aria che entra attraverso gli stomi della foglia.

Invece, nella « C_4 » *A. rosea*, le cellule che circondano le sottili vene della foglia sono disposte in modo del tutto differente; esse formano due cilindri concentrici alla venatura. Molte cellule nelle foglie delle piante « C_4 » contengono pochi cloroplasti e probabilmente non danno un contributo significativo al processo fotosintetico, ma le cellule che appartengono ai due cilindri hanno un altissimo numero di cloroplasti. Il cilindro piú esterno è costituito da quelle che sono chiamate cellule del mesofillo. Le cellule del cilindro piú interno, noto come « rivestimento del fascio » hanno pareti sottili.

Abbiamo scoperto che gli enzimi fotosintetici contenuti nelle cellule del mesofillo sono differenti dagli enzimi delle cellule del rivestimento del fascio. Questo significa che i due strati di cellule hanno funzioni diverse. Per esempio è soltanto nelle cellule del mesofillo che l'anidride carbonica atmosferica viene fissata dal ciclo « C_4 ». Allo stesso modo è soltanto nelle cellule del rivestimento del fascio che i prodotti del ciclo « C_4 », acido malico e acido aspartico, vengono scissi per rilasciare l'anidride carbonica fissata e contemporaneamente per formare acido piruvico.

Il passaggio dei composti « C_4 » dai siti del mesofillo dove avviene la fissazione del carbonio ai siti del rivestimento del fascio dove essi sono scissi, e il passaggio di ritorno dell'acido piruvico dalle cellule del rivestimento del fascio alle cellule del mesofillo sono due processi che effettuano un trasporto netto di anidride carbonica dallo strato esterno a quello piú interno. Non appena l'anidride carbonica diffonde nelle cellule del rivestimento del fascio, viene fissata e successivamente convertita in zucchero e in altri composti finali della fotosintesi con il ciclo Calvin-Benson. Dal momento che i siti per la fissazione del carbonio nel ciclo di Calvin-Benson sono localizzati all'interno delle cellule del rivestimento del fascio, non usufruiscono di un contatto diretto con l'anidride carbonica atmosferica. Se le cellule piú

L'istogramma mostra che l'eredità di un'efficiente fotosintesi « C_4 » è complessa e non è stata raggiunta dall'incrocio delle piante affini « C_3 » e « C_4 ». Le coppie di sbarrette rappresentano l'andamento fotosintetico di *A. rosea*, di un ibrido di prima generazione e di *A. patula* in atmosfera normale (*in grigio*) e in atmosfera con poco ossigeno (*in colore*). L'ossigeno inibisce la fotosintesi nelle « C_3 » e nell'ibrido ma non nelle « C_4 ».

interne non fossero rifornite dall'anidride carbonica, prodotta col ciclo « C_4 » dalle cellule più esterne, esse non potrebbero sintetizzare nessun prodotto fotosintetico utile e le piante « C_4 » non crescerebbero.

Il vantaggio del sistema fotosintetico biciclico delle piante « C_4 » è che esse possono assorbire in modo molto efficiente anidride carbonica atmosferica poco concentrata e immettere l'anidride carbonica nel ciclo Calvin-Benson per ottenere i prodotti finali della fotosintesi necessari per la crescita. L'anatomia delle foglie specializzate fornisce poi la compartimentazione spaziale richiesta dalle due distinte catene di reazione ed è inoltre una componente essenziale per la fotosintesi delle piante « C_4 ». Il sistema biciclico logicamente richiede un'entrata maggiore di energia per molecola di carbonio fissata, ma questo extra è bilanciato da un'alta velocità di fissazione di anidride carbonica alle basse concentrazioni che sono caratteristiche di tutti gli spazi intercellulari delle foglie sia di piante « C_4 » che di piante « C_3 ».

È geneticamente determinato che una pianta possegga l'uno o l'altro sistema fotosintetico. Quale è allora la probabilità che in un ceppo di piante « C_3 » possa emergere un sistema « C_4 »? Esperimenti di ibridazione con *A. rosea* e *A. petula* sono stati condotti nel nostro laboratorio; il nostro collega Malcolm A. Nobs è stato in grado di ibridizzare *A. rosea* con altre specie « C_3 » di *Atriplex*. Questi sono i soli incroci raggiunti finora anche se un gran numero di piante « C_3 » e « C_4 » dello stesso genere sono conosciute. I risultati degli esperimenti con l'*Atriplex* sono illuminanti.

Gli ibridi di prima generazione, prodotti dalla fecondazione delle piante di *A. rosea* con il polline di *A. patula*, sono sotto molti punti di vista intermedi tra i due genitori, ma nella generazione seguente appare una fisionomia più caratteristica. Dal punto di vista dell'anatomia della foglia, per esempio, le foglie degli ibridi di prima generazione hanno una morfologia intermedia, e le foglie delle piante di seconda generazione hanno caratteri morfologici che ricordano quelli di uno o dell'altro genitore. Una simile segregazione di caratteri è molto evidente se guardiamo le caratteristiche biochimiche delle piante; un esempio è offerto dalla considerevole variabilità della quantità di PEP-carbossilasi presentata dalle foglie delle piante di seconda generazione.

Pochi ibridi di seconda generazione assomigliano al loro genitore « C_4 » sia nell'anatomia della foglia che nelle caratteristiche biochimiche. Uno potrebbe aspettarsi che gli ibridi posseggano anche un funzionale sistema fotosintetico del ciclo a quattro atomi di carbonio. Ma delle parecchie centinaia di piante finora esaminate questo non si è verificato. Al contrario tutti gli ibridi di sintetizzavano a velocità molto simili. Non abbiano trovato una pianta che possedesse una completa integrazione del sistema « C_4 ». I nostri risultati suggeriscono che un'efficiente fotosintesi a doppio ciclo richiede che non solo siano presenti tutti gli elementi necessari, ma anche che questi siano perfettamente e completamente coordinati. Così anche se solo pochi geni possono essere implicati nel determinare l'eredità genetica di ogni componente della fotosintesi « C_4 », l'assoluta necessità di una completa coordinazione delle proprietà anatomiche e biochimiche della foglia potrebbe rendere del tutto impossibile che questo processo si manifesti nelle piante che ne sono prive. Poiché è evidente che il sistema « C_4 » si è evoluto indipendentemente un certo numero di volte, probabilmente i nostri risultati riflettono anche la notevole differenza effettiva tra i processi naturali e lo sforzo dell'uomo con manipolazioni genetiche artificiali.

A nostro parere è a disposizione dell'uomo un metodo per utilizzare il processo « C_4 » che è di gran lunga più promettente dei tentativi di introdurre il sistema stesso nelle piante che ne sono attualmente prive. Il granturco, la canna da zucchero, il sorgo e certe erbe da pascolo già possiedono questo processo fotosintetico. Questo gruppo di piante potrebbe essere aumentato incrociando alcune specie selvatiche di piante « C_4 » con piante domestiche utili dal punto di vista agricolo. L'amaranto (affine alla *Tidestromia*), una delle principali piante coltivate anticamente nell'America centrale, è un esempio della specie « C_4 » che potrebbe essere incrociata per produrre sementi utili per il foraggio animale.

Per quanto riguarda le attuali specie da raccolto « C_4 », un problema importante concerne la modalità del loro impiego. Per i nostri studi sulle specie selvatiche come la *Tidestromia* e le due specie di *Atriplex*, noi siamo sicuri del fatto che questo meccanismo rappresenti un adattamento agli ambienti caldi e aridi. Più precisamente c'è una componente che determina un incremento di fotosintesi ad alte temperature e un aumento di rendimento in caso di efficiente utilizzazione dell'acqua durante la crescita. Sono in corso di studio tecniche agricole che possano trarre pieno vantaggio, sfruttando queste potenzialità? Non crediamo che ve ne siano. I programmi di incroci e le applicazioni pratiche hanno avuto come principale scopo il raggiungimento del massimo rendimento possibile in condizioni ottimali. Sebbene questo obiettivo possa essere valido per le campagne più ricche, è invece di valore discutibile nelle regioni produttivamente svantaggiose, dove per esempio il costo o la disponibilità di una risorsa come l'acqua limita la quantità di terra coltivabile.

Per i loro attributi speciali le piante domestiche « C_4 » potrebbero rivelarsi le coltivazioni ideali per queste terre.

GIOCHI MATEMATICI

di Martin Gardner

Ancora sul libero arbitrio, con un'affascinante paradosso di William Newcomb relativo alla predizione

Secondo l'opinione comune, la controversia sul libero arbitrio ha ormai esaurito tutti gli argomenti disponibili, e nessun nuovo paladino può far altro che rifriggere argomenti stantii e noti a chiunque. È questo un grave errore. Io non conosco alcun argomento che abbia possibilità migliori di aprire nuove vie.

WILLIAM JAMES

Uno tra i problemi eterni della filosofia è come dare una spiegazione (o render ragione) della natura del libero arbitrio. Se il concetto viene spiegato nell'ambito di un modo di pensare deterministico, la volontà cesserà di esser libera, quale che sia il senso che si dà comunemente a questo vocabolo, ed è difficile vedere in qual modo si potrà sfuggire al fatalismo. È inutile cercare di sottrarsi al proprio destino. Perché lavorar duro per procurare un futuro migliore a se stessi o ad altri se quel che si fa è sempre quel che non si potrebbe comunque non fare? E come si potrà rimproverare a qualcuno un certo comportamento, se egli non avrebbe potuto fare altrimenti?

Ugualmente inutili sembrano d'altra parte i tentativi di spiegare la volontà in un ambito di pensiero indeterministico. Se un'azione non è causata da uno stato precedente di noi stessi e del mondo, è difficile vedere come si possa impedire all'azione di essere puramente casuale. La nozione secondo cui le decisioni verrebbero prese nella mente in base a una qualche sorta di meccanismo casuale non aiuta molto a risolvere il dilemma se la volontà si debba considerare libera o no.

I filosofi non hanno mai raggiunto l'accordo sul modo in cui evitare i corni di questo dilemma. Anche all'interno di una singola scuola non sono mancati netti disaccordi. William James e John Dewey, due tra i principali esponenti del pragmatismo americano, ci forniscono un ottimo esempio in proposito. Pur essendo Dewey un tenace difensore delle libertà democratiche, la sua metafisica considerava il comportamento umano come totalmente determinato da quella che James chiamava la « spinta del passato ». Per Dewey il libero arbitrio è altrettanto illusorio di quanto lo è nella psicologia di B. F. Skinner. James era al contrario un indeterminista radicale. Secondo lui la mente umana avrebbe il potere di introdurre nella storia qualcosa di genuinamente nuovo, e neppure Dio stesso sarebbe in grado di conoscere il futuro se non in parte. « *Questo* », egli scrisse, « è ciò che conferisce la sua palpitante realtà alla nostra vita morale e che la fa fremere... di un'eccitazione così strana e complessa ».

Una terza impostazione, approfondita da Immanuel Kant, accetta entrambe le soluzioni come modi ugualmente veri ma incommensurabili di considerare il comportamento umano. Per Kant la situazione è un po' simile a quella illustrata in uno dei *grooks* di Piet Hein:

Pensando qualche volta
all'enigma della vita,
simile parmi questa
a due cassette chiuse, e la chiave
che potrebbe aprire una
all'interno dell'altra è contenuta.

Il libero arbitrio non è né destino né caso. In un qualche modo insondabile esso partecipa di entrambi. Ciascuno è la chiave che apre l'altro. Non è, questo, un concetto contraddittorio, come quello di un triangolo quadrato, ma un paradosso che l'esperienza ci impone e la cui soluzione trascende il pensiero umano. In questo modo lo vide anche Niels Bohr. Egli trovò questa situazione simile al suo « principio di complementarità » nella meccanica quantistica. È questo un punto di vista che Einstein, che era uno spinoziano, considerò disgustoso mentre molti altri fisici, e tra questi basti citare Oppenheimer, lo trovarono molto attraente.

Ma che cos'ha a che fare il libero arbitrio con i giochi matematici? La risposta è che, negli ultimi decenni, i filosofi della scienza si sono rotti la testa con un gran numero di strani « paradossi della predizione » aventi a che fare col problema della volontà. Alcuni di questi paradossi possono essere considerati nel modo migliore come situazioni di gioco. Si elabora una matrice dei pagamenti e si cerca di determinare qual è la migliore strategia che un giocatore deve seguire, ma si finisce col trovarsi impigliati in un labirinto di ambiguità sconcertanti concernenti il tempo e la causalità.

Un esempio sorprendente di tale paradosso venne alla luce nel 1970 in un articolo, *Newcomb's Problem and Two Principles of Choice* di Robert Nozick, filosofo alla Harvard University. Il paradosso è così profondo, così divertente, così affascinante, con i pensatori così equamente divisi in due campi l'un contro l'altro armato che esso promette di produrre una letteratura più vasta di quella che si occupa del paradosso della predizione dell'impiccagione inattesa.

Il paradosso di Newcomb si chiama così da colui che lo escogitò, William A. Newcomb, fisico teorico al Lawrence Livermore Laboratory dell'Università della California. (Suo nonno era fratello di Simon Newcomb, lo astronomo.) Newcomb immaginò il problema nel 1960, mentre meditava su un famoso paradosso della teoria dei giochi chiamato il « dilemma del prigioniero ». Pochi anni dopo, del problema di Newcomb venne a conoscenza Nozick attraverso il comune amico Martin David Kruskal, matematico alla Princeton University. « Non so se ho il diritto di presentare questo scritto » osserva Nozick. « Si tratta di un problema magnifico. Vorrei averlo inventato io ». Pur non riuscendo a risolverlo, Nozick decise di metterlo in ogni modo per iscritto. Il suo articolo apparve in *Essays in Honor of Carl G. Hempel*, a cura di Nicholas Rescher, pubblicati da D. Reidel nel 1970. Quanto segue è in gran parte una parafrasi dell'articolo di Nozick.

		ESSERE	
		MOSSA 1 (PREDICE CHE NOI PRENDIAMO SOLO LA SCATOLA 2)	MOSSA 2 (PREDICE CHE PRENDIAMO ENTRAMBE LE SCATOLE)
NOI	MOSSA 1 (PRENDIAMO SOLTANTO LA SCATOLA 2)	$ 1 000 000	$ 0
	MOSSA 2 (PRENDIAMO ENTRAMBE LE SCATOLE)	$ 1 001 000	$ 1000

La matrice dei pagamenti per il paradosso di Newcomb.

Su un tavolo ci sono due scatole chiuse, S_1 e S_2. S_1 contiene 1000 dollari, S_2 non contiene nulla oppure un milione di dollari. Le due scatole non presentano alcun segno esterno che possa servire come indicazione del contenuto. La scelta è tra due azioni:

1) Prendere il contenuto di entrambe le scatole.
2) Prendere solo il contenuto di S_2.

In un tempo indeterminato prima del test, un essere superiore ha compiuto una predizione su ciò che decideremo di fare. Non è necessario adottare una concezione deterministica, ma soltanto essere convinti del fatto che le predizioni di tale Essere sono « quasi certamente » corrette. Volendo possiamo identificare tale Essere con Dio, ma il paradosso conserva il suo significato anche se lo identifichiamo con un'intelligenza superiore proveniente da un altro pianeta oppure con un super calcolatore in grado di sondare il nostro cervello e di fare predizioni molto precise su quelle che saranno le nostre decisioni. Se l'Essere si attende che noi scegliamo entrambe le scatole, avrà lasciato vuota S_2. Se si aspetta che scegliamo solo S_2, avrà messo in essa un milione. (Se egli si attende che noi prendiamo la nostra decisione a caso, per esempio gettando in aria una moneta, avrà lasciato S_2 vuota.) S_1 contiene in ogni caso 1000 dollari. Noi comprendiamo perfettamente la situazione, l'Essere sa che la comprendiamo, noi che egli sa e così via.

Che cosa dovremmo fare? È chiaro che non è un vantaggio per noi affidarci alla monetina, e perciò dovremo decidere con la nostra testa. Il paradosso risiede nel fatto imbarazzante che c'è un buon argomento a favore di ciascuna decisione; i due argomenti non possono però essere giusti entrambi. Il problema consiste allora nello spiegare perché uno dei due è sbagliato.

Consideriamo dapprima l'argomento a favore della scelta della sola S_2. Crediamo che l'Essere sia dotato di eccellenti doti di predizione. Se prendiamo entrambe le scatole, l'Essere avrà quasi certamente anticipato la nostra azione e avrà lasciato S_2 vuota. In tal caso avremo soltanto i 1000 dollari contenuti in S_1. Al contrario, se si prende soltanto S_2 l'Essere, attendendosi tale decisione, avrà messo quasi certamente in essa un milione di dollari. È perciò chiaramente a nostro vantaggio prendere soltanto S_2.

Quest'argomentazione è convincente? Sì, ma supponiamo che l'Essere abbia fatto questa predizione diciamo una settimana fa e poi non se ne sia più occupato. I casi quindi sono due: o egli avrà messo un milione di dollari in S_2, oppure l'avrà lasciata vuota. « Se in S_2 c'è il denaro, ci rimarrà quale che sia la nostra scelta. Esso non può sparire. Se invece non c'è, non vi apparirà all'improvviso se sceglieremo di prendere solo il contenuto della seconda scatola ». Si suppone che non operi alcuna « causalità all'indietro », ossia che le nostre azioni presenti non possano influenzare ciò che l'Essere ha fatto la scorsa settimana. Perché allora non prendere entrambe le scatole ed entrare quindi in possesso del contenuto di entrambe? Se S_2 contiene il milione, otterremo 1 001 000 dollari. Se è vuota, avremo almeno 1000 dollari. Se saremo così pazzi da scegliere soltanto S_2, sappiamo che non potremo avere più di un milione di dollari, e c'è anche una possibilità, per quanto remota, di rimanere a mani vuote. È perciò chiaramente a nostro vantaggio prendere entrambe le scatole!

« Ho posto questo problema a numerose persone, amici e studenti in classe », scrive Nozick. « A quasi tutti il da farsi apparve perfettamente chiaro e ovvio. La difficoltà consiste nel fatto che queste persone sembrano dividersi quasi equamente in due campi opposti, e ciascuna delle due parti pensa che l'altra sia del tutto stupida ».

« Dati due argomenti opposti così stringenti, non ci si può accontentare della propria convinzione di non aver dubbi sul da farsi, né ci si può limitare a ripetere semplicemente una delle due argomentazioni. Si deve anche disarmare l'argomentazione contraria, distruggere la sua forza pur dimostrandole il dovuto rispetto ».

Nozick enuclea come segue la « forza » dei due argomenti. Supponiamo che l'esperimento sia già stato fatto prima molte volte. In tutti i casi l'Essere ha fatto la predizione esatta. Coloro che hanno scelto entrambe le scatole hanno avuto sempre solo 1000 dollari, mentre chi ha scelto solo S_2 ha sempre un milione di dollari. Non abbiamo ragione di supporre che il nostro caso sarà diverso. Se un amico stesse osservando la scena, sarebbe per lui del tutto razionale scommettere, pagando anche poste molto alte, che se scegliamo entrambe le scatole avremo solo 1000 dollari. Di fatto, se dopo la nostra scelta delle due scatole ci fosse un intervallo di tempo prima dell'apertura delle stesse sarebbe ragionevole scommettere anche per noi, offrendo poste molto alte, che avremo solo 1000 dollari. Conoscendo ciò, non saremmo pazzi a scegliere entrambe le scatole?

Purtroppo l'altro argomento dimostra che non siamo meno pazzi se non lo facciamo. Supponiamo che S_1 sia trasparente. Vediamo naturalmente che essa contiene 1000 dollari. Non possiamo invece vedere all'interno di S_2, ma il lato opposto della scatola è trasparente ed è visibile al nostro amico, che siede di fronte a noi. Egli sa quindi se la scatola è vuota o se contiene un milione di dollari. Benché egli non dica nulla, comprendiamo che, sia S_2 piena o vuota, egli desidera che noi scegliamo entrambe le scatole. Egli lo desidera perché indipendentemente dal contenuto di S_2, tale scelta rappresenterà 1000 dollari in più. Perché non trarre vantaggio dal fatto che l'Essere ha mosso per primo e non può modificare la sua mossa?

Nozick, che è uno specialista nella teoria della decisione, affronta il paradosso considerando analoghe situazioni di gioco in cui, come qui, sussista un conflitto tra due principi di scelta in apparenza ugualmente validi: il « principio dell'utilità attesa » e il « principio di dominanza ». Per vedere in che modo il principio si applichi, consideriamo la matrice dei pagamenti per il gioco di Newcomb (*si veda la figura nella pagina a fronte*). L'argomentazione a favore della scelta della sola S_2 deriva dal principio che si dovrebbe scegliere in modo da rendere massima l'utilità attesa (il valore per noi) del risultato. La teoria dei giochi calcola la utilità attesa di ogni azione moltiplicando ciascuno dei suoi risultati reciprocamente esclusivi per la probabilità del risultato. Abbiamo supposto che l'Essere predica con una certezza quasi assoluta, ma usiamo una certa prudenza e consideriamo la probabilità pari a 0,9. L'utilità attesa di scegliere entrambe le scatole è di (0,1 × $ 1 001 000) + (0,9 × $ 1000) = = $ 101 000. L'utilità attesa di prendere solo S_2 è (0,9 × $ 1 000 000) + (0,1 × × $ 0) = $ 900 000. Se ci facciamo guidare da questo principio, la nostra strategia migliore consiste nello scegliere soltanto la seconda scatola.

Il principio della dominanza è però intuitivamente altrettanto valido. Supponiamo che il mondo sia diviso in n situazioni diverse. Per ciascuna situazione k ci si offrono azioni reciprocamente esclusive. Se in almeno una situazione ci troviamo in vantaggio scegliendo a, e in tutte le altre a è la scelta migliore oppure le scelte sono

equivalenti, il principio di dominanza dice che si dovrebbe scegliere *a*. Consideriamo ancora la matrice delle poste in palio nella pagina a fronte. Le situazioni sono i risultati delle due mosse dell'Essere. La soluzione consistente nel prendere entrambe le scatole è fortemente dominante. In ogni situazione essa ci dà infatti 1000 dollari in piú di quelli che otterremmo prendendo soltanto la seconda scatola.

Fino a questo punto possiamo giungere seguendo l'analisi di Nozick, ma i lettori interessati dovrebbero perfezionarla per quanto concerne le sue imbarazzanti situazioni di conflitto relative al problema di Newcomb. Nozick perviene infine alle seguenti conclusioni provvisorie:

Se crediamo in un determinismo assoluto e riteniamo che l'Essere abbia veramente previsto il nostro comportamento con assoluta precisione, dovremmo « scegliere » (qualunque cosa ciò possa significare!) di prendere soltanto S_2. Supponiamo per esempio che l'Essere sia Dio e che noi siamo calvinisti devoti, fermamente convinti che Dio conosca ogni benché minimo particolare del nostro futuro. Oppure supponiamo che l'Essere possegga un congegno in grado di viaggiare nel tempo, congegno che egli può lanciare nel futuro e riportare indietro con un film di ciò che avremo fatto in quell'occasione futura quando avremo operato la nostra scelta. Credendo ciò, noi dovremmo prendere soltanto S_2, fermamente convinti che la sensazione di aver compiuto veramente una scelta sia una mera illusione.

Nozick ci ricorda nondimeno che il paradosso di Newcomb *non* suppone che l'Essere abbia un potere di predizione perfetto. Se crediamo di possedere una particella per quanto piccola di libero arbitrio (o alternativamente che qualche volta l'Essere possa sbagliare, diciamo una volta ogni 20 miliardi di casi), allora questa potrebbe essere una delle volte in cui l'Essere ha sbagliato. La nostra decisione piú saggia sarà allora quella di prendere entrambe le scatole.

Nozick non è soddisfatto di questa conclusione. « La differenza tra uno in *n* e niente in *n*, per un *n* finito grande a piacere, potrebbe giustificare questa differenza? E in che modo il fatto che l'Essere è certo della sua predizione potrebbe dissolvere la forza dell'argomento della dominanza? » Entrambe le domande rimangono senza risposta. Nozick spera che la pubblicazione del problema « possa suscitare una soluzione che mi consenta di non perdere piú il mio tempo a meditare su di esso ».

Una tale soluzione, « per ridare la serenità a Nozick », fu tentata da Maya Bar-Hillel e da Avishai Margalit dell'Università ebraica di Gerusalemme nel loro articolo *Newcomb's Paradox Revisited* apparso nel « British Journal for the Philosophy of Science ». Essi adottano la stessa impostazione propria della teoria dei giochi usata da Nozick ma pervengono alla conclusione opposta. Anche se l'Essere non è in grado di fare predizioni assolutamente certe, essi raccomandano di prendere soltanto la seconda scatola. Dobbiamo rassegnarci al fatto, essi sostengono, che la nostra migliore strategia consiste nel comportarci *come se* l'Essere avesse fatto una predizione corretta, pur sapendo che c'è una piccola possibilità che abbia errato. Sappiamo che ha mosso prima di noi, ma non possiamo giocare meglio se gli toccasse muovere dopo di noi. « Non possiamo infatti superare l'Essere se non sapendo che cosa ha predetto, ma non possiamo sapere, e neppure congetturare fondatamente che cosa ha predetto prima di compiere realmente la nostra scelta ».

Possiamo aver l'impressione, scrivono Bar-Hillel e Margalit, che sia operante una causalità all'indietro – che la nostra scelta renda in qualche modo piú probabile la presenza nella seconda scatola del milione di dollari – ma son tutte fandonie. Noi scegliamo soltanto S_2 in base all'argomentazione che la causalità è indipendente dalla direzione del tempo – e che quindi la nostra decisione è realmente la « causa » del fatto che la seconda scatola sia vuota o contenga il milione di dollari – Newcomb propose la seguente variante del suo paradosso. Entrambe le scatole sono trasparenti. S_1 contiene i soliti 1000 dollari. S_2 contiene un pezzo di carta con su scritto un numero intero abbastanza grande. Non sappiamo se il numero è un numero primo o se invece è composto. Se esso risulterà un numero primo (naturalmente non dobbiamo verificare questo fatto prima di fare la nostra scelta) riceveremo un milione di dollari. L'Essere avrà optato per un numero primo se avrà predetto la scelta, da parte nostra, della sola scatola S_2, e per un numero composto se avrà previsto la scelta, da parte nostra, di entrambe le scatole.

È chiaro che un atto di volontà da parte nostra non riuscirà a trasformare il numero da un numero primo a un numero composto e viceversa. La natura del numero è fissata per l'eternità. Perché allora non prendere entrambe le scatole? Se il numero è primo, otterremo 1 001 000 dollari. Se non lo è, avremo almeno 1000 dollari. (Incernente un fatto matematico determinabile, che non può però essere accertato se non una volta compiuta la scelta.)

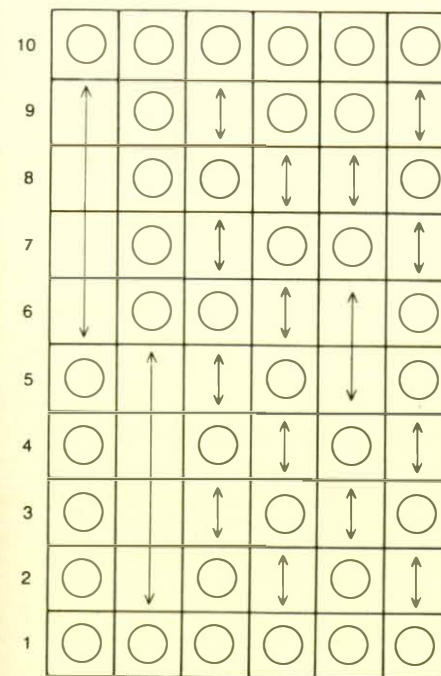
È facile escogitare altre varianti. Per esempio, ci sono 100 scatolette contenente ciascuna un biglietto da 10 dollari. Se l'Essere si attende che le prendiamo tutte, non mette in esse nient'altro, mentre se si attende che ne prendiamo una sola – forse scelta a caso – avrà aggiunto in quella scatoletta un grande diamante. Supponiamo che siano già state fatte migliaia di prove, e di aver partecipato come giocatori a una metà di esse. Ogni volta, con pochissime eccezioni, il giocatore che ha preso una sola scatola ha trovato in essa il diamante, mentre il giocatore che ha preso tutte le scatole vi ha trovato solo il denaro. Agendo in modo pragmatico, sulla base dell'esperienza passata, dovremmo prendere una scatola sola. Ma allora, come rifiutare la logica dell'argomentazione che dice che avremo tutto da quadagnare e nulla da perdere se la prossima volta prenderemo tutte le scatole?

Queste varianti non aggiungono nulla di sostanzialmente nuovo. Con riferimento alla versione originale, Nozick raccomanda, sia pure con qualche esitazione, di prendere entrambe le scatole. Bar-Hillel e Margalit ci consigliano vivamente di « aggregarci al club dei milionari » scegliendo solo S_2. Questa è anche l'opinione di Kruskal e di Newcomb. Ma ciascuna delle due parti ha veramente fatto qualcosa di piú che non ripetere semplicemente la propria opinione? Può essere che il paradosso di Newcomb legittimi il libero arbitrio invalidando la possibilità di massima di un Essere, quale che esso sia, capace di indovinare la scelta di una persona tra due azioni ugualmente razionali con una precisione maggiore del 50 per cento?

Che cosa ne pensa il lettore?

Applicato al problema dell'ascensore l'algoritmo di Richard M. Karp mostra che occorrono almeno 72 viaggi per spostare le 45 persone nell'edificio di 9 piani. La soluzione al secondo problema dell'ascensore, che si trova nel volume di Knuth citato il mese scorso, richiede le seguenti 12 tappe:

1. 123456 in 2.
2. 112334 in 3.
3. 224456 in 4.
4. 135566 in 5.
5. 112334 in 6.
6. 224456 in 2.
7. 222444 in 4.
8. 222222 in 2.
9. 555666 in 5.
10. 666666 in 6.
11. 111333 in 3.
12. 111111 in 1.

La soluzione al problema dell'ascensore giapponese, che è una versione raffinata di un vecchio problema combinatorio, è mostrata nella figura qui sotto.

Una risposta al problema dell'ascensore.

Parecchi lettori ci hanno scritto protestando perché nel numero di agosto, proponendo il problema del π, non abbiamo specificato anche la possibilità di usare l'elevamento a potenza. A tutti i lettori dobbiamo una scusa però, per quanto matematici, sempre di giochi si tratta! E inoltre, tirando un po' la corda, l'elevamento a potenza può essere fatto rientrare nella moltiplicazione.

Ma a parte l'arrabbiatura, molti lettori ci hanno mandato un sacco di soluzioni diverse da quelle da noi proposte. Alcune erano diverse solo formalmente, nel senso che usavano lo stesso numero di π ma in forma diversa, altre erano invece decisamente migliori di quelle da noi proposte, cioè riuscivano a esprimere gli interi con un numero minore di π.

Per esempio il signor Bartolomucci di San Donato ci ha proposto $15 = [(\pi + \sqrt{\pi}) \times \pi]$ e $19 = [(\pi + \pi) \times \pi]$. Quest'ultima soluzione ci è stata proposta anche dal signor Lodato di Roma che ha proposto anche $20 = [\pi^\pi : \sqrt{\pi}]$, usando però la potenza! Parecchi altri lettori ci hanno inviato soluzioni piú o meno simili, ma non possiamo citarli tutti: abbiamo citato soltanto i primi arrivati. Comunque ringraziamo tutti.